

Dottorato di Ricerca in Studi Giuridici Comparati ed Europei

Ciclo XXXVIII

Curriculum in diritto amministrativo, costituzionale e internazionale

**Artificial intelligence in law enforcement: a critical analysis of
the EU AI Act's framework for protecting public values and
fundamental rights**

Dottoranda: Marianna Merler

Supervisor: Prof.ssa Barbara Marchetti

Co-supervisor: Prof. Simone Penasa

Coordinatore: Prof. Federico Puppo

ANNO ACCADEMICO 2024-2025



UNIVERSITÀ DEGLI STUDI DI TRENTO

FACOLTÀ DI GIURISPRUDENZA
Dottorato in Studi Giuridici
Comparati ed Europei

Candidata: Marianna Merler

*Artificial intelligence in law enforcement:
a critical analysis of the EU AI Act's framework
for protecting public values and fundamental rights*

Supervisor: Prof.ssa Barbara Marchetti

Co-supervisor: Prof. Simone Penasa

Anno Accademico 2024-2025

Indirizzo specialistico in Diritto amministrativo, costituzionale e internazionale

XXXVIII ciclo

Esame finale: 24 aprile 2026

Commissione esaminatrice:

Prof.ssa Luisa Torchia, Università degli Studi Roma Tre

Prof. Andrea Simoncini, Università degli Studi di Firenze

Prof.ssa Nicoletta Rangone, Università LUMSA

TABLE OF CONTENTS

TABLE OF CONTENTS.....	7
ABSTRACT.....	11
INTRODUCTION	13
1. Context and motivation.....	13
2. Research problem and significance	15
3. Methodology and scope	16
4. Structure of the Thesis and contribution to the debate	18
Theoretical coordinates on artificial intelligence and law enforcement	23
1. Overview of the Chapter	23
Part I: Theoretical coordinates on artificial intelligence	25
1. A brief history of AI	25
1.1. Before the birth of AI.....	25
1.2. The birth of AI: The Summer Seminar at Dartmouth College.....	30
1.3. Between “AI springs” and “AI winters”	31
1.4. The present of AI	34
1.5. The future of AI	37
1.6. Different ideas of AI	38
2. Technical coordinates on artificial intelligence	42
2.1. Model building.....	45
2.2. Inference phase	50
3. From causation to correlation	52
4. A legal definition of AI.....	54
5. An overview of the AI lifecycle.....	62
Part II: Theoretical coordinates on law enforcement.....	66
1. The complex and fragmented landscape of law enforcement in the European Union	66
1.1. Defining criminal offence, criminal penalty and public security.....	69
1.2. National security and public security.....	74
2. Delimiting the concept of “law enforcement authorities” under the European Union Framework	81
3. Law enforcement and human rights: striking a delicate balance	88
CHAPTER 2	95
Artificial intelligence enters law enforcement	95
1. Overview of the Chapter	95

2.	The digitisation of police organisations	99
Part I: Predictive policing		105
1.	The current frontier of data analytics: predictive policing.....	105
1.1.	The experience of predictive policing in the United States	105
1.2.	Defining predictive policing	109
1.3.	Predictive policing in Europe.....	119
2.	Considerations on predictive policing	126
2.1.	Data and data protection concerns	129
2.2.	Bias and non-discrimination	133
2.3.	Errors and accuracy concerns	139
2.4.	Transparency issues	140
Part II: Automated video surveillance and face recognition technology.....		143
1.	The police use of video analytics.....	143
1.1.	Image processing through Automated Video Surveillance	144
1.2.	Identifying individuals through biometric technologies	151
2.	An example of hard biometrics: face recognition technology	153
2.1.	Face recognition technology: how does it work?	155
2.2.	The use of face recognition technology: where does Europe stand?	162
3.	Considerations on Automated Video Surveillance and Face Recognition Technology	164
3.1.	Data and data protection concerns	164
3.2.	Errors, accuracy concerns and bias	170
3.3.	The ever-closer relationship between public authorities and private entities	176
3.	Function creep in algorithmic policing.....	181
CHAPTER 3		189
The regulation of AI systems for law enforcement purposes within the AI Act		189
1.	Overview of the Chapter.....	189
Part I: the Artificial Intelligence Act		192
1.	The European path towards the regulation of AI.....	192
2.	The first binding rules on AI: the European Union's Artificial Intelligence Act	194
3.	The complex legal architecture of the AI Act.....	205
3.1.	The horizontal approach	206
3.2.	The risk-based approach	209
3.3.	The by-design regulation approach.....	218
3.4.	The co-regulatory approach: the technical standards.....	222
3.5.	Considerations on the overall AI Act's structure	228

Part II: Law enforcement in the AI Act.....	232
1. The AI Act’s provisions in the law enforcement context	232
2. Predictive Policing in the AI Act	232
2.1. Prohibited predictive policing.....	236
2.2. Other categories of predictive policing.....	242
3. Automated video surveillance in the AI Act.....	245
3.1. AI systems for categorising natural persons	246
3.2. Biometric identification systems.....	250
3.3. Biometric verification	261
4. Considerations on the established legal framework.....	263
CHAPTER 4	271
Transparency and fairness in AI-driven law enforcement: a critical analysis under the EU AI Act.....	271
1. Overview of the Chapter	271
Part I: Transparency issues in the AI Act	274
1. Transparency in AI systems.....	274
2. The multiple dimensions of transparency in the AI Act	277
2.1. The first dimension of transparency: information obligations.....	278
2.2. The second dimension of transparency: traceability.....	282
2.3. The third dimension of transparency: interpretability	283
2.4. The fourth dimension of transparency: explainability	287
3. Considerations on the established legal framework with a focus on law enforcement	289
3.1. Technical barriers.....	297
3.2. Technological barriers.....	300
3.3. Human oversight.....	307
3.4. The tactical barriers and the difficulty in ensuring the right to an effective remedy	313
Part II: Algorithmic fairness under the AI Act: implications for law enforcement.....	322
1. Algorithmic fairness.....	322
2. The legal principle of non-discrimination.....	327
3. The AI Act’s provisions on bias and non-discrimination	329
3.1. Discrimination as a potential risk of using AI systems: the central role of a Fundamental Rights Impact Assessment	330
3.2. From data to discrimination: how the AI Act confronts bias in AI-powered law enforcement.....	339
3.3. The role of human oversight in mitigating algorithmic discrimination	350
CONCLUDING REMARKS.....	357

1. A final assessment: is the AI Act sufficiently adequate in protecting rights and public values?	358
2. The delicate relationship between AI and personal data	365
3. Insights for human rights law	371
4. Organisational changes for responsible AI innovation	373
4.1. A proper organisational culture	374
4.2. People	376
4.3. Processes	378
BIBLIOGRAPHY	383
Laws and Regulations	448
Supranational	448
National	449
Case Law	451
Supranational Courts	451
National Courts	453

ABSTRACT

The increasing deployment of artificial intelligence (hereinafter “AI”) systems by law enforcement authorities across the European Union has intensified long-standing concerns regarding the protection of fundamental rights, the rule of law, transparency, and democratic accountability. While AI-based technologies promise enhanced efficiency, predictive capacity, and resource optimisation in policing, their use in this particularly sensitive domain risks exacerbating discrimination, opacity, and arbitrary decision-making. These risks directly implicate the core values enshrined in Article 2 TEU and demand a robust and coherent regulatory response.

Against this backdrop, the present Thesis examines whether the existing European legal framework, most notably Regulation (EU) 2024/1689, is adequate to ensure that AI systems deployed by law enforcement authorities operate lawfully, fairly, and in full compliance with fundamental rights and public values. To that end, the analysis critically engages with the AI Act’s regulatory architecture as applied to two of the most contested uses of AI in policing: predictive policing and automated video surveillance.

To address the research question, the Thesis investigates three interrelated issues: first, whether the AI Act’s definition of “AI system” adequately captures law enforcement systems that pose significant risks to fundamental rights; second, whether the Regulation’s concept of “law enforcement” is sufficiently clear to ensure legal certainty and consistent risk classification; and third, whether the AI Act, in conjunction with EU data protection law, provides substantive protection beyond minimum standards, or instead leaves structural gaps in the governance of algorithmic policing. Methodologically, the Thesis adopts a doctrinal legal analysis, complemented by interdisciplinary insights from computer science, criminology, and ethics, which are necessary to understand both the technological object of regulation and the policing strategies that inform its use.

The Thesis argues that while the AI Act represents a significant step forward in the EU’s approach to AI governance, its application to law enforcement remains marked by conceptual ambiguities, regulatory fragmentation, and enforcement challenges. These shortcomings risk undermining the effective protection of fundamental rights and public values unless addressed through careful interpretation, complementary safeguards, and a principled application of transparency and fairness requirements tailored to the law enforcement context.

INTRODUCTION

1. Context and motivation

The proliferation of artificial intelligence (hereinafter “AI”) across public and private sectors has fundamentally altered the landscape of service delivery and decision-making within the European Union. Among the manifold applications of AI, its deployment in law enforcement contexts stands out as particularly consequential and contested. Police authorities collect vast amounts of data daily, storing them in large databases that require continuous analysis to inform operational and strategic decisions. It is widely acknowledged that digital technologies – of which AI represents the current frontier – can support law enforcement authorities in extracting meaningful insights and achieving a more comprehensive understanding of the information they gather, thereby enhancing the effectiveness and efficiency of their functions.

Law enforcement authorities increasingly rely on algorithmic tools to predict crime hotspots, identify suspects through biometric recognition and assess recidivism risk. These applications promise enhanced efficiency, data-driven objectivity, and optimised resource allocation in the prevention, investigation, detection, and prosecution of criminal offences, as well as in the safeguarding of public security.

The potential benefits are substantial: AI systems can process volumes of data that would overwhelm human analysts, identify patterns invisible to conventional investigation, and enable more targeted deployment of limited law enforcement resources. Yet the integration of AI into policing also raises profound challenges to foundational principles of the European legal order. AI systems deployed in law enforcement operate at the intersection of state power and individual liberty, a domain already characterised by structural asymmetries, coercive authority, and far-reaching consequences for personal autonomy, dignity, and equality. The opacity of algorithmic decision-making processes, the risk of embedding or amplifying discriminatory patterns, and the potential for pervasive surveillance threaten core values enshrined in Article 2 of the Treaty on the European Union: respect for human dignity, freedom, democracy, equality, the rule of law, and fundamental rights. When public authorities delegate or augment their decision-making power through automated systems, traditional safeguards of legality, transparency, and accountability face unprecedented strain. These concerns have fuelled an intense debate regarding the appropriateness of delegating critical decisions to AI systems, with both

INTRODUCTION

academic and political spheres emphasising the necessity of developing robust regulatory frameworks capable of ensuring ethical and responsible use of these technologies.

These tensions have catalysed intensive regulatory activity at the European level. Although European institutions began taking the first steps towards AI regulation as early as 2021, a definitive regulation, the Artificial Intelligence Act (hereinafter “AI Act”), entered into force on 1 August 2024, establishing the world’s first comprehensive horizontal regulatory framework for AI systems¹. The AI Act adopts a risk-based approach, imposing differentiated obligations depending on the level of risk posed by specific AI applications. It distinguishes between prohibited uses, high-risk, limited-risk, and minimal-risk applications, establishing obligations that vary according to the level of risk that AI systems pose to health, safety, and fundamental rights. Law enforcement applications occupy a privileged position within this regulatory architecture: certain uses are categorically prohibited under Article 5, while most others are classified as high-risk systems subject to stringent requirements and obligations. This regulatory treatment reflects legislative recognition that AI in law enforcement implicates especially sensitive fundamental rights and demands heightened scrutiny.

The Italian context provides instructive examples of the challenges associated with AI deployment in law enforcement. Several AI-enhanced video surveillance projects in experimental phases have been blocked by the Italian Data Protection Authority, which, after requesting and obtaining information regarding technical specifications, purposes, and legal bases for data processing, deemed them contrary to existing personal data protection legislation. These cases share common deficiencies: the absence of adequate legal bases and the lack of transparency regarding the concrete operation of these systems, the data upon which they were trained, their accuracy, and their potential for discriminatory outcomes. Despite this ambitious regulatory intervention, significant questions remain regarding the AI Act’s capacity to ensure rights-compliant, transparent, and fair deployment of AI systems by law enforcement authorities. The Regulation’s effectiveness depends critically on how key concepts are defined and interpreted, how regulatory obligations are operationalised in practice, and how the AI Act interacts with parallel legal frameworks.

As I will analyse, the Regulation introduces a highly intricate set of classifications, obligations, and compliance mechanisms and, due to its complexity, effective implementation requires substantial organisational adaptation. To ensure that AI systems are used lawfully, transparently, and in accordance

¹ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

with the AI Act's requirements, law enforcement authorities will need to establish new governance structures or revise existing ones, as well as develop the necessary expertise and operational routines.

2. Research problem and significance

The deployment of AI systems in law enforcement presents a regulatory trilemma. On one hand, law enforcement authorities face legitimate pressures to modernise their investigative capacities, optimise limited resources, and leverage technological innovation to enhance public safety. On the other hand, the use of automated or semi-automated systems in this domain threatens to undermine long-established safeguards that protect individuals from arbitrary state power. Additionally, regulatory intervention itself risks being either too restrictive – stifling beneficial innovation – or too permissive – failing to prevent harms to fundamental rights.

The AI Act's risk-based regulatory model attempts to navigate this trilemma by calibrating obligations to the severity and probability of harm. Yet this approach presupposes clear definitional boundaries and reliable risk classification mechanisms. To address this trilemma, this Thesis is structured around one central research question and three subsidiary research questions that guide the analysis.

Central research question:

Are the existing legal and regulatory frameworks sufficient to ensure the effective protection of fundamental rights and public values when law enforcement authorities carry out their functions using AI systems?

Subsidiary research questions:

SRQ 1: *Is the AI Act's definition of "AI system" sufficiently clear; and comprehensive to capture all relevant systems, including those at the margins of the regulatory perimeter; and to provide legal certainty to both providers and deployers operating in this domain?*

SRQ 2: *Is the AI Act's notion of "law enforcement" sufficiently precise to determine, with reasonable certainty, when a given activity falls within its scope? What are the consequences of any definitional ambiguity for the identification of applicable obligations, the classification of risk levels, and the allocation of regulatory responsibilities among the actors involved?*

SRQ 3: *Does the AI Act's risk-classification framework adequately reflect the diversity and complexity of AI systems deployed for law enforcement purposes? Furthermore, is the relationship between the AI Act and the applicable data protection legislation — particularly the Law Enforcement Directive —*

INTRODUCTION

sufficiently coherent and well-coordinated to afford robust and consistent protection of fundamental rights in practice?

These questions are not merely academic. How they are resolved will determine whether European law enforcement authorities can deploy AI systems in ways that genuinely respect fundamental rights, or whether regulatory frameworks provide only symbolic protection while allowing practices that erode transparency, reproduce discrimination, and concentrate unchecked power.

The significance of this research extends beyond doctrinal legal analysis. First, it contributes to ensuring that regulatory frameworks align with constitutional principles and human rights obligations that bind EU Member States. Second, it provides practical guidance to law enforcement authorities, technology providers, and oversight bodies seeking to interpret and apply complex and sometimes ambiguous legal requirements. Third, it informs ongoing policy debates about the future evolution of AI governance, both within Europe and in jurisdictions worldwide that look to the EU as a regulatory model. Addressing the compatibility of AI-enhanced policing with applicable legal frameworks is essential not only to ensure compliance with legal obligations but also to foster public trust and social acceptance, prerequisites for law enforcement authorities to fulfil their mandate effectively and maintain democratic legitimacy.

3. Methodology and scope

This Thesis employs a predominantly doctrinal legal methodology, centred upon the systematic analysis, interpretation, and critical evaluation of legal texts, including primary EU legislation, case law from the Court of Justice of the European Union (CJEU) and the European Court of Human Rights (ECtHR), and relevant secondary sources, which encompass Commission guidelines, opinions of the European Data Protection Board, decisions of national Data Protection Authorities, policy documents and reports issued by supranational and international bodies. The research is grounded in extensive engagement with scholarly contributions from the academic literature, which constitute a substantial component of the analytical framework employed throughout this investigation.

The doctrinal analysis is complemented by interdisciplinary insights drawn from computer science, criminology, and ethics. Understanding the regulatory object requires a sufficient grasp of how AI systems function technically, including AI systems' architectures, training data dependencies, and algorithmic decision-making processes. The AI Act itself presupposes knowledge of the technology it regulates in order to interpret and evaluate its concrete applicability, presenting itself as legislation that

INTRODUCTION

must be equally sustainable from both legal and technological perspectives. Similarly, analysing the deployment context requires familiarity with policing strategies and criminological theories that underlie the approaches adopted by public authorities to enhance effectiveness and efficiency, primarily through optimal resource allocation and the shift from reactive to preventive policing strategies. Finally, the AI Act's regulatory choices reflect ethical principles articulated in soft law instruments produced by the High-Level Expert Group on AI (HLEG), the Organisation for Economic Co-operation and Development (OECD), United Nations Interregional Crime and Justice Research Institute (UNICRI), International Criminal Police Organization (INTERPOL), and the Council of Europe. Understanding these ethical foundations is essential to comprehending the normative substrate that has oriented European regulation in this field. Engaging with these interdisciplinary dimensions is essential to contextualising legal analysis, yet the interpretative framework remains fundamentally juridical. Therefore, the research remains legal research, and engagement with other fields of knowledge is instrumental to legal analysis.

The research activity has been enriched through research periods conducted at institutions providing complementary perspectives and expertise. A research period at Tallinn University of Technology (TalTech), with its interdisciplinary environment, facilitated the acquisition of necessary technical knowledge and competencies. Additionally, research conducted at the General Directorate of the municipality of Trento provided valuable practical insights. At the commencement of this research, the municipality was a partner in two European Union-funded projects (MARVEL and Protector) directed toward the experimentation of AI-enhanced video surveillance systems to improve public security policies. Following the opening of an investigation by the Italian Privacy Authority, involvement in these proceedings enabled access to materials that proved instructive for understanding the practical and regulatory challenges confronting public authorities deploying AI systems.

Scope limitations: This Thesis focuses on AI systems used by or on behalf of law enforcement authorities within the European Union. It does not comprehensively address AI in judicial decision-making, immigration control, or national security contexts, except where such applications intersect with law enforcement functions. The analysis concentrates on two specific use cases – predictive policing and automated video surveillance – as paradigmatic examples that illuminate broader regulatory challenges, rather than attempting an exhaustive survey of all law enforcement AI applications. These applications exemplify the regulatory challenges at stake and serve as concrete case studies through which to assess the AI Act's conceptual coherence, practical applicability, and normative adequacy.

INTRODUCTION

Geographically, the research centres on EU law but incorporates comparative references to national implementations, particularly where Member States have adopted divergent approaches or where national case law illuminates interpretative questions. The research does not constitute a systematic and structured comparison between multiple legal systems; rather, it examines selected jurisdictions on specific aspects where comparative analysis yields valuable insights. Attention is devoted to European experiences. Instructive examples may be drawn from Germany, particularly the judgment of the Bundesverfassungsgericht of 16 February 2023 concerning the hessenDATA platform, and from France, including Law No. 2023-380 of 19 May 2023 and Conseil constitutionnel decision No. 2023-850 DC of 17 May 2023 regarding AI-enhanced video surveillance during the Olympic Games. Examining these jurisdictions enables critical engagement with regulatory approaches and provides insights toward potential improvements in the legal framework.

4. Structure of the Thesis and contribution to the debate

The Thesis comprises four substantive Chapters followed by Concluding remarks:

Chapter 1 establishes foundational concepts and definitional frameworks. Part I provides technical background on AI, tracing its historical evolution from symbolic AI through machine learning to contemporary deep learning systems, a synthetic overview of the principal characteristics of AI systems and critically examines the AI Act's definition of "AI system" in Article 3(1). These technical foundations are necessary to delimit the boundaries of the AI concept and to examine in depth those aspects of the technology whose legal implications are investigated in later Chapters. Part II analyses the concept of "law enforcement", exploring its legal boundaries through EU primary and secondary law, relevant jurisprudence, and the specific articulation of law enforcement purposes in the AI Act and data protection instruments. This dual foundation is necessary to delimit the regulatory object (what counts as AI) and the regulatory domain (what counts as law enforcement use).

Chapter 2 examines the intersection of AI and law enforcement through two critical applications. Part I addresses predictive policing, analysing data-driven techniques for forecasting crime locations, times, and potential offenders or victims. It examines the epistemological assumptions underlying predictive models, their operational deployment in European jurisdictions, and the fundamental rights implications. Part II focuses on AI-enabled video surveillance and face recognition technologies, analysing their technical capabilities, current deployments and impacts on privacy, data protection, freedom of assembly, and equality. The Chapter demonstrates that while these technologies offer certain operational benefits and have justified their adoption on grounds of improved efficiency and

INTRODUCTION

resource allocation, they also generate systemic risks to fundamental rights that demand robust regulatory safeguards.

Chapter 3 provides a comprehensive analysis of the AI Act's regulatory architecture as applied to law enforcement contexts. Part I explicates the Regulation's horizontal scope, risk-based approach, design-oriented obligations, and co-regulatory reliance on harmonised standards. Part II examines how the AI Act categorises AI systems for law enforcement purposes across its regulatory tiers: prohibited practices (Article 5), high-risk systems (Article 6 and Annex III), and limited- or minimal-risk AI systems. The analysis identifies which specific law enforcement applications fall within each category, what obligations attach, and where regulatory gaps or ambiguities arise. Particular attention is paid to interpretative challenges surrounding the scope of prohibitions, the boundaries of the high-risk classification, and the adequacy of conformity assessment procedures.

Chapter 4 concentrates on two cross-cutting normative dimensions essential for ensuring public trust and acceptance in this sensitive domain: transparency and fairness. Part I examines transparency requirements across the AI value chain, including documentation obligations for providers, information duties toward deployers and affected persons, and the complex relationship between transparency and technical explainability. It assesses whether the AI Act's transparency provisions are adequate for law enforcement contexts, where both operational secrecy concerns and democratic accountability imperatives create distinctive tensions. Part II addresses fairness, analysing both discrimination risks embedded in training data and algorithmic design, and broader concerns about procedural justice, equity, and legitimacy. Issues of bias and discriminatory outcomes are particularly evident in law enforcement contexts, as illustrated by cases such as COMPAS. The Chapter evaluates whether the AI Act's fairness-related provisions – read alongside EU anti-discrimination and data protection law – provide sufficient safeguards, or whether structural gaps remain.

Finally, Concluding remarks draw together the findings of the analysis and provide a final assessment addressing the central research question directly. This section examines whether the AI Act is sufficiently adequate in protecting rights and public values, explores the delicate relationship between AI and personal data processing, considers insights from human rights law regarding the importance of legal bases that meet stringent standards of clarity, specificity, and proportionality, and discusses the organisational changes necessary for responsible AI innovation within law enforcement structures.

This Thesis makes several original contributions to the emerging scholarship on AI regulation and law enforcement:

First, it provides conceptual clarification. More specifically, it provides systematic analysis of two foundational concepts – “AI system” and “law enforcement” – whose interaction determines the AI

INTRODUCTION

Act's scope and applicability. The analysis demonstrates how uncertainty surrounding these definitions affects risk classification and the determination of applicable obligations.

Second, it provides application-specific analysis. By focusing on predictive policing and face recognition/AI-enabled video surveillance as concrete case studies, the Thesis bridges the gap between abstract regulatory principles and operational realities. It examines specific systems deployed across European jurisdictions, including SARI Real-Time in Italy, hessenDATA in Germany, augmented cameras during the French Olympic Games, and various municipal initiatives (such as XLAW, Keycrime, and Giove for predictive policing, and experimental face recognition projects in Como and Turin). The analysis demonstrates how general provisions function (or fail to function) in specific law enforcement contexts and examines how national Data Protection Authorities and Courts have responded to these deployments, highlighting common deficiencies such as the absence of adequate legal bases and insufficient transparency.

Third, it provides an integrative doctrinal approach. The Thesis synthesises the AI Act with EU data protection law (GDPR and LED), fundamental rights jurisprudence from the CJEU and the ECtHR, as well as judgments from national Courts, and soft law instruments to present a holistic account of the regulatory constellation governing AI systems in law enforcement, an integration rarely attempted in existing scholarship.

Fourth, it provides a critical normative assessment: Beyond descriptive analysis, the Thesis offers normative evaluation of the AI Act's adequacy in protecting transparency, fairness, and fundamental rights, identifying specific regulatory gaps and proposing interpretative strategies to strengthen protection. It examines both the strengths of the regulatory framework (such as the prohibition of certain practices and the detailed high-risk regime) and its limitations (including challenges in risk classification, implementation difficulties, and potential for function creep whereby technologies initially introduced under narrow conditions gradually expand into broader applications).

Fifth, the analysis has interdisciplinary grounding. By incorporating technical and criminological perspectives while maintaining a juridical analytical frame, the Thesis models how legal scholarship can engage productively with AI governance challenges that transcend traditional disciplinary boundaries. It addresses not only legal requirements but also the organisational dimensions of responsible AI innovation, including the need for appropriate organisational culture, workforce expertise (AI literacy), and processes for risk management throughout the AI lifecycle.

Sixth, it provides practical implementation insights. The Thesis provides concrete guidance on implementation challenges, examining the importance of legal bases that meet stringent standards of clarity, specificity, and proportionality; the need for law enforcement authorities to develop new governance structures and operational routines; the critical role of stakeholder engagement and public

INTRODUCTION

trust; and the necessity of coordinated intervention across organisational culture, human capital, and processes to support responsible AI innovation.

Ultimately, this research aims to contribute to ensuring that as law enforcement authorities increasingly rely on AI systems to carry out their functions, European legal frameworks evolve to guarantee that such reliance remains compatible with constitutional principles, fundamental rights, and democratic values that define the European legal order. The Thesis demonstrates that while the AI Act provides an essential regulatory foundation and represents a significant advancement, effective protection of public values requires careful attention to definitional clarity, risk classification accuracy, implementation capacity, and the interaction between multiple overlapping regulatory regimes. Although the AI Act establishes important baseline protections, it frequently defers critical substantive questions to harmonised standards, national implementing measures, or parallel regulatory regimes, creating a complex regulatory landscape whose effective navigation demands sustained institutional commitment and ongoing critical scrutiny.

CHAPTER 1

Theoretical coordinates on artificial intelligence and law enforcement

1. Overview of the Chapter	23
Part I: Theoretical coordinates on artificial intelligence	25
1. A brief history of AI	25
1.1. Before the birth of AI	25
1.2. The birth of AI: The Summer Seminar at Dartmouth College	30
1.3. Between “AI springs” and “AI winters”	31
1.4. The present of AI.....	34
1.5. The future of AI.....	37
1.6. Different ideas of AI.....	38
2. Technical coordinates on artificial intelligence	42
2.1. Model building.....	45
2.1.1. Knowledge representation	45
2.1.2. Machine learning	46
2.1.3. Deep learning.....	48
2.2. Inference phase.....	50
3. From causation to correlation	52
4. A legal definition of AI.....	54
5. An overview of the AI lifecycle	62
Part II: Theoretical coordinates on law enforcement	66
1. The complex and fragmented landscape of law enforcement in the European Union	66
1.1. Defining criminal offence, criminal penalty and public security	69
1.2. National security and public security	74
2. Delimiting the concept of “law enforcement authorities” under the European Union Framework	81
3. Law enforcement and human rights: striking a delicate balance	88

1. Overview of the Chapter

The first Chapter introduces the two conceptual pillars whose intersection constitutes the central analytical focus of this Thesis: artificial intelligence (hereinafter, AI) and law enforcement.

The first part of Chapter 1 (Part I) establishes the necessary technical foundations for the subsequent legal analysis. It reconstructs the historical development of AI, identifies its principal technological trajectories, and examines the core components and operational logics of contemporary AI systems.

CHAPTER 1

Emphasis is placed on the definition of AI adopted by the EU Artificial Intelligence Act (hereinafter “AI Act”), as this definition is determinative of the Regulation’s material scope of application. A rigorous legal assessment of AI-based systems cannot be conducted in abstraction from their technological characteristics. Indeed, the AI Act itself presupposes a sufficient understanding of the technological artefacts it regulates, as such understanding is indispensable for delineating its scope, interpreting its normative provisions, and assessing their concrete applicability. This methodological and epistemic interdependence between technical and legal analysis reflects the inherently interdisciplinary nature of AI regulation and informs the overall structure of the Thesis, which seeks to systematically relate the technical properties of AI systems to the legal regimes governing their deployment.

The second part of the Chapter (Part II) turns to the concept of “law enforcement” and examines its normative contours within the EU legal order. It analyses the legal boundaries of law enforcement by reference to EU primary and secondary legislation, relevant jurisprudence of the Court of Justice of the European Union, and the specific formulation of law enforcement purposes in the AI Act and in EU data protection instruments. This inquiry serves to clarify the functional scope and legal meaning of law enforcement activities, thereby providing an essential analytical framework for evaluating the use of AI-based technologies in this domain.

As will be demonstrated throughout the subsequent Chapters, it is the intersection between AI and law enforcement that generates particularly acute ethical and legal challenges. The deployment of AI systems in law enforcement contexts raises fundamental questions regarding the adequacy of existing safeguards, the protection of fundamental rights, and the preservation of core public values within an increasingly automated and data-driven regulatory environment.

Part I: Theoretical coordinates on artificial intelligence

1. A brief history of AI

Summarising in a few pages the history of AI is far from simple, and it risks being reductive and incomplete. However, since the present Thesis is centred around the use of AI by law enforcement authorities, it is thus necessary to provide at least a footnote of history of this complex technology, highlighting the triumphs and failures, the concepts and ideas that have mostly contributed to the changes that in a few decades have led to previously unimaginable achievements that are still a long way from a definitive point of arrival.

1.1. Before the birth of AI

Throughout human history, there has been a persistent quest to create entities that possess intelligence beyond our own¹. This fascination is reflected in various narratives, from the legend of the Golem of Prague, as recounted in *Magic Prague*², to Mary Shelley's *Frankenstein*³. The concept of intelligent machines can also be traced back to the ancient tales of automata in civilizations such as Egypt, China, and Greece⁴, later in the play of Karel Čapek⁵ and finally in the works of modern science fiction writers like Arthur C. Clarke⁶ and Isaac Asimov⁷.

Beyond the narratives, over the centuries various studies have given important contributions to the creation of intelligent entities, eventually culminating in the AI known today. The roots of these studies can be traced in different disciplines, decades or centuries old. Philosophy provided important insights into the later construction of intelligent systems. Aristotle's formulation of the laws governing the mind and his study on syllogism⁸ represent the birth of logical formalization, which allows one to

¹ G. LIVELEY, S. THOMAS, *Homer's Intelligent Machines: AI in Antiquity*, in S. CAVE, K. DIHAL, S. DILLON (Eds.), *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, Oxford Academic, 2020, 21-42.

² A.M. RIPELLINO, *Magic Prague*, University of California Press, 1993.

³ M.W. SHELLEY, *Frankenstein or, The Modern Prometheus*, Modern Library, 1993.

⁴ In ancient Greek reference goes to the myth of Pygmalion and Galatea and the legend of Talos.

⁵ K. ČAPEK, *R.U.R. (Rossum's Universal Robots)*, Miraggi Edizioni, 2022.

⁶ A.C. CLARKE, *2001: A Space Odyssey*, Perfection Learning, 2000.

⁷ I. ASIMOV, *The Complete Robots*, Harper Voyager, 2018. I. ASIMOV, *Runaround*, in *I, Robot*, Mondadori, 2021. In this novel the author formulated the Three Laws of Robotics: 1) A robot may not injure a human being or, through inaction, allow a human being to come to harm; 2) A robot must obey orders given it by human beings except where such orders would conflict with the First Law; 3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Law. Some decades later, Asimov added a Fourth, or Zeroth Law, which precedes the others: 0) A robot may not harm humanity, or, by inaction, allow humanity to come to harm. I. ASIMOV, *Robots and Empire*, Harper Voyager, 2023.

⁸ ARISTOTLE, *Prior analytics, Book I*, Oxford University Press, 2009.

generate conclusions given initial valid premises. Much later, Thomas Hobbes proposed that reasoning is like a numerical computation⁹ and Gottfried Wilhelm Leibniz also advanced the idea of reasoning as a calculable technique, *i.e.* one that can be performed by means of numerical calculations¹⁰. Later studies focused on the automation of computation: the Pascaline, built by Blaise Pascal, represents the most famous early calculating machine, capable of adding and subtracting¹¹ and Leibniz went further building a calculator that could execute more operations than Pascal's (add, subtract and multiply)¹². The idea underpinning these studies was that the mind operates, at least in part, according to logical rules and that it was possible to build physical systems that emulate some of those rules. The philosophical discourse then focused on the question of whether the mind itself *is* a physical system¹³, what the source of knowledge is¹⁴ and what the link between knowledge and action is. Centuries of philosophical studies that have sought to answer these dilemmas have thus built the premises of core AI's theoretical ideas and concepts. Moreover, philosophy is still relevant in AI-related studies, fostering new questions and new perspectives on critical aspects of current AI¹⁵.

In relation to the representation of knowledge, mathematical studies became extremely relevant. Indeed, these studies, particularly those connected to logic, computation and probability, became necessary to move from theory to practice¹⁶. In the seventeenth century, Leibniz envisioned a universal and formal language, the *characteristica universalis*, capable of expressing mathematical, scientific, and metaphysical concepts through a series of symbols¹⁷. In his idea this language could have been

⁹ W. MOLESWORTH, *The Collected Works Of Thomas Hobbes*, Routledge, 1997.

¹⁰ G.W. LEIBNIZ, *The New Method of Learning and Teaching Jurisprudence*, Lawbook Exchange Ltd, 2017.

¹¹ A mechanical calculator was designed by Leonardo Da Vinci, but he never built it. The first (known) calculating machine was constructed around 1623 by Wilhelm Schickard. However, the Pascaline invented by Blaise Pascal is the most famous. «The arithmetical machine produces effects which approach nearer to thought than all the actions of animals. But it does nothing which would enable us to attribute will to it, as to the animals». B. PASCAL, *Pascal's Pensées*, E. P. Dutton & Co., Inc., 1958, 96.

¹² Both the Pascaline and the Leibniz's Machine were automatic calculators. However, they were mechanical, not electronic, and they were not programmable.

¹³ Different philosophical approaches can be identified: rationalism, dualism and materialism. René Descartes is a strong advocate of the philosophy of rationalism and dualism: he believed in the power of reasoning in understanding the world, but he also suggested that the mind cannot be governed *entirely* by physical laws, otherwise there would be no space for free will. There is a part of the mind thus that is exempt from physical laws.

¹⁴ The question «Where does knowledge come from?» gave rise to the empiricism movement and the logical positivism.

¹⁵ Ethics, as a branch of philosophy, had a fundamental role in AI research and developments, especially in the last decade. Among others, see J. MORLEY, L. FLORIDI, L. KINSEY, A. ELHALAL, *From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices in Science and Engineering Ethics*, 26, 2020, 2141-2168. S. VALLOR, B. GREEN, I. RAICU, *Overview of Ethics in Tech Practice*, Project Ethics in Technology Practice of Marrkula Center for Applied Ethics, June 2018. Available at: <https://www.scu.edu/ethics-in-technology-practice/overview-of-ethics-in-tech-practice/>. S. VALLOR, B. GREEN, I. RAICU, *Technology and Engineering Practice: Ethical Lenses to Look Through*, Project Ethics in Technology Practice of Marrkula Center for Applied Ethics, July 2020. Available at: <https://www.scu.edu/ethics-in-technology-practice/ethical-lenses/>. L. FLORIDI, *Soft Ethics and the Governance of the Digital in Philosophy & Technology*, 31, 2018.

¹⁶ S. RUSSELL, P. NORVIG, *Artificial Intelligence. A modern approach*, Pearson Education Inc., 2010, 7.

¹⁷ Leibniz introduced his first idea of a universal formal language in his work *Dissertatio de arte combinatoria*, but he further developed it in later studies. G.W. LEIBNIZ, *Dissertatio de arte combinatoria*, 1666.

used as part of a universal logical calculus, named *calculus ratiocinator*. Later, in his work *Explication de l'Arithmétique Binaire*¹⁸ he explained the modern binary system¹⁹, which represents the basis for binary code used in computer programming²⁰. In the nineteenth century, George Boole described an algebraic system, known as Boolean algebra, which is based on the binary system and on three basic operators, namely AND, OR and NOT²¹. Friedrich Ludwig Gottlob Frege extended Boole's logic including objects and relations, introducing the second-order logic.

Based on these studies, Kurt Gödel went further and presented the two incompleteness Theorems which showed limits on deduction, a fundamental idea in modern logic, showing that there are functions that cannot be computed²². These Theorems formed the basis of Alan Turing and Alonzo Church's work which culminated in the Thesis that any algorithmic problem, namely a problem for which there is a precise solution, can be solved using simple machines, meaning machines that can perform just a few elementary operations²³. Contemporarily, Turing conceived a conceptual model of a universal machine, known as the Turing Machine, that is capable of simulating any algorithm²⁴. The Turing Machine is still relevant today since modern calculators match the concept of universal machines. Gödel's contribution also led Turing to investigate further what problems admit an algorithmic solution – and thus can be deemed computable – and those that do not²⁵.

In addition to the notion of computability, the notion of tractability had a huge impact on later developments²⁶. The problem of tractability is closely related to the time it takes to solve a problem

¹⁸ G.W. LEIBNIZ, *Explication de l'Arithmétique Binaire*, 1703.

¹⁹ See M.R. ANTOGNAZZA, *Leibniz: An Intellectual Biography*, Cambridge University Press, 2009, 357.

²⁰ The binary system represents all the numbers with only two digits: 0 and 1. The binary code represents the particular digital representation system used in automatic information processing. A bit (derived from binary digit) is the minimum unit of information processed by a calculator and it can assume two possible values: 0 and 1. All information, even the most complex, can be described as a sequence of 0 and 1: it is sufficient to establish a correspondence between a *bitmap* and what one wants to represent (e.g. letters of the alphabet, a picture, a sound). In informatics, the use of the binary system, rather than the decimal system, is due to the ability of physical components – and in particular electronic components – to assume only two alternative “positions”: on/off, positive/negative. Moreover, a parallelism between binary system and logic is possible: just as a digit can only take on two values, 1 and 0, so a proposition can only take on two alternative values, true and false.

²¹ G. BOOLE, *The Mathematical Analysis of Logic*, Philosophical Library New York, 1948. Boolean algebra represents the basis of modern symbolic logic. Indeed, Boole realised that combinations between propositions follow precise rules, similar to those governing arithmetic operations. These rules determine the functioning of the logical operators AND, OR and NOT, the conjunctions that serve to link propositions in more complex structures. Claude Shannon was the first to realise electronic circuits applying the Boolean algebra.

²² *Gödel's Incompleteness Theorems*, Stanford Encyclopedia of Philosophy. First published Nov 11, 2013 with substantive revision in April 2020. Available at: <https://plato.stanford.edu/entries/goedel-incompleteness/>.

²³ G. SARTOR, *L'informatica giuridica e le tecnologie dell'informazione. Corso d'informatica giuridica*, Giappichelli, 2022, 47.

²⁴ A.M. TURING, *On Computable Numbers, with an Application to the Entscheidungsproblem*, in *Proceedings of the London Mathematical Society*, s2-42(1), 1937, 230–265. See also *Turing Machines*, Stanford Encyclopedia of Philosophy. First published Sep 24, 2018. Available at: <https://plato.stanford.edu/entries/turing-machine/>.

²⁵ For further discussion on the topic see SARTOR, *L'informatica giuridica e le tecnologie dell'informazione*, cited, 48.

²⁶ «Roughly speaking, a problem is called intractable if the time required to solve instances of the problem grows exponentially with the size of the instances». RUSSELL, NORVIG, *Artificial Intelligence. A modern approach*, cited, 8.

and investigates which problems can only be solved in exponential time and can therefore be considered intractable. A method on how to recognise intractable problems was provided by the theory of NP-completeness of Stephen Cook and Richard Karp²⁷. Another fundamental theory in AI-related studies is the theory of probability, initially framed by Gerolamo Cardano and further developed by other researchers. Among them, a key role is acknowledged to Thomas Bayes, whose work underpins many important approaches in AI systems. Indeed, probability theories are at the essence of current AI developments.

Studies within other disciplines are relevant to the emergence of AI as a new formal science. Some interesting insights into the development of intelligent agents have been drawn from economics, from utility theory, decision theory and game theory²⁸. The idea of “rational agents”, *i.e.* agents who make decisions in such a way as to maximise gains, is strictly connected to these economic studies. Furthermore, studies in neuroscience, a discipline which is focused on how the brain works and how it processes information for making decisions and taking actions, also played a fundamental role. Advances in this discipline demonstrated that the brain structure consisted of a network of cells, called neurons, which transmitted electro-chemical impulses. The first theories thus sought to explain how thoughts, actions and consciousness were the result of signals transmitted between neurons in the human brain. Based on this new knowledge, Warren S. McCulloch and Walter Pitts presented the first mathematical model of neural activity in 1943²⁹. Moreover, psychology, especially cognitive psychology, which looks at the brain as an information-processing device³⁰, helped in answering the question on *how* humans think and act and linguistics focused on the relationship between thoughts and language.

Computer engineering had a fundamental role in the realization of the artifact, the material part of an intelligent system. Interestingly, one of the most influential programmable machines was a loom, devised by Joseph Marie Jacquard. In the nineteenth century Charles Babbage, inspired by Jacquard’s loom, designed two machines, the Difference Engine and the Analytical Engine³¹, but he never completed them. Contemporarily, Augusta Ada King, Countess of Lovelace wrote programs for the unfinished Babbage’s analytical engine, becoming the first programmer in history. The first modern calculators, digital and electronic, were developed during the World War II. In Great Britain, under

²⁷ *Computational Complexity Theory*, Stanford Encyclopedia of Philosophy. First published the 27 July 2015 with substantive revision July 2016. Available at: <https://plato.stanford.edu/entries/computational-complexity/>.

²⁸ J. VON NEUMANN, O. MORGENSTERN, *The Theory of Games and Economic Behavior*, Princeton University Press, 2007.

²⁹ W.S. MCCULLOCH, W. PITTS, *A logical calculus of ideas immanent in nervous activity* in *Bulletin of Mathematical Biophysics*, 5, 1943, 115-133. Based on the studies of McCulloch and Pitts, Hebb went further building a neural network capable of learning and in 1958 Rosenblatt created the Perceptron.

³⁰ W.H.F. BARNES, *The Nature of Explanation* in *Nature*, 3890, 1944, 605.

³¹ For a description of the Analytical Engine see A.G. BROMLEY, *Charles Babbage’s Analytical Engine, 1838* in *IEEE Annals of the History of Computing*, 20(4), 1998, 29-45.

CHAPTER 1

Newman's direction Heath Robinson was designed and built, an operational computer for deciphering war messages of the German enemies. Some years later, the same group developed Colossus, a programmable electronic digital computer. In Germany, Konrad Zuse developed the Z3, which represents the first programmable computer based on the binary system. In the United States, John Atanasoff and Clifford Berry assembled the Atanasoff Berry Computer (ABC), often cited as the first electronic computer, and the University of Pennsylvania presented the Electronic Numerical Integrator and Computer (ENIAC), electronic and universal, but still based on the decimal system. The technological step towards modern computers was the von Neumann Machine, developed by John von Neumann, which was able to internally memorize the programme³² and execute instructions sequentially. The architecture of von Neumann Machine was adopted by the Electronic Discrete Variable Automatic Computer (EDVAC), which inspired later producers.

Until today, there has been rapid evolution and radical changes, but the main characteristics of computers have remained almost unchanged³³: computers are digital, which means that they process information expressed in the digits 0 and 1; computers are electronic, exploiting the behaviour of electrons; computers are programmable, which means that they operate by following the instructions of computer programmes, which express algorithms; computers are universal, which means that any computer can, at least in principle, execute any algorithm.

The advances in control theory and cybernetics also played an important role, further developing studies on self-controlling machines, notably machines that do not need the intervention of a human entity in their functioning process³⁴. Apart from philosophy, mathematics, economics, neuroscience, psychology, computer engineering, control theory and cybernetics and linguistics³⁵, some authors suggest that other disciplines might be added, especially in the light of current developments in the field³⁶.

³² The von Neumann machine had a control unit (CU) and an arithmetic-logic unit (ALU) that together make up the Central Processing Unit (CPU), also called the processor. The control unit identifies the instructions to be followed and the relevant data and the ALU executes the instructions. The data and instructions are accessed from the central memory. The CPU can roughly be described as a complex combination of 'switches', activated by the calculator itself, that change their position during execution, allowing or disallowing the flow of electricity (on and off/1 and 0). For further discussion see SARTOR, *L'informatica giuridica e le tecnologie dell'informazione*, cited, 49-50.

³³ *Ibidem*, 66.

³⁴ A. ROSENBLUETH, N. WIENER, J. BIGELOW, *Behavior, Purpose and Teleology*, *Philosophy of Science*, 10(1), 1943, 18–24. Available at: <http://www.jstor.org/stable/184878>. The authors introduced the term *cybernetics*. A few years later, Wiener defined the scope and goals of the new discipline in N. WIENER, *Cybernetics: Or Control and Communication in the Animal and the Machine*, MIT Press, 1961.

³⁵ These are the disciplines that Russell and Norvig have described as being crucial in AI developments. See RUSSELL, NORVIG, *Artificial Intelligence. A modern approach*, cited, 5-16.

³⁶ G.M. CAMPEDELLI, *Machine learning for criminology and crime research. At the crossroads*, Routledge, 2022, 24.

1.2. The birth of AI: The Summer Seminar at Dartmouth College

All these studies led to enormous progress, but they tended to remain within their respective disciplines, which at the same time tended not to communicate with others. The turning point occurred in the mid-twentieth century. Indeed, there is unanimous agreement in the scientific community that AI as a distinct field of study has a precise date of birth: 1956, the year of the Summer Seminar at Dartmouth College in Hanover, New Hampshire.

A few months earlier, in August 1955, John McCarthy from Dartmouth College, Marvin Minsky from Harvard University, Claude E. Shannon from Bell Telephone Laboratories and Nathaniel Rochester from IBM Corporation, submitted a proposal to the Rockefeller Foundation stating their goal of bringing together a group of researchers interested in building machines capable to perform intelligent tasks³⁷. At the conference, some of the brightest minds in the field joined the original group³⁸.

After six weeks spent together, the concrete results of the Dartmouth Conference fell short of expectations and the goals originally set were only partially achieved³⁹. Despite this, the event remained the catalyst that allowed a group of scientists from different domains to come together and discuss what would later become an immense undertaking of research and technological innovation⁴⁰. Indeed, the Dartmouth conference was accompanied by a climate of widespread enthusiasm and

³⁷ J. MCCARTHY, M. MINSKY, C. SHANNON, N. ROCHESTER, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955. Available at: <http://www-formal.stanford.edu/jmc/history/dartmouth.pdf> <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>. “We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer”. The aspects on which they wanted to focus were: 1) Automatic computers, 2) How can a computer be programmed to use a language, 3) Neuron Nets, 4) Theory of the size of a calculation, 5) Self-improvement, 6) Abstractions, 7) Randomness and creativity. “They had in common a belief (more like a faith at that point) that what we call thinking could indeed take place outside the human cranium, that it could be understood in a formal and scientific way, and that the best nonhuman instrument for doing it was the digital computer”. P. MCCORDUCK, *Machines who think. A Personal Inquiry into the History and Prospects of Artificial Intelligence*, Routledge, 2004, 111.

³⁸ The researchers who joined the conference were Allen Newell and Herbert Simon from Carnegie Tech, Arthur Samuel from IBM, Ray Solomonoff and Oliver Selfridge from MIT, Trenchard More from Princeton.

³⁹ In MCCORDUCK, *Machines who think*, cited, 114, the author quotes McCarthy’s words: “Anybody who was there was pretty stubborn about pursuing the ideas that he had before he came, nor was there, as far as I could see, any real exchange of ideas. People came for different periods of time (...) so not everybody was there at once. It was a great disappointment to me because it really meant that we couldn’t have regular meetings”.

⁴⁰ In 2006 a conference was held in Dartmouth of which Moor leaves us an interesting testimony. One of the main objectives of the conference was to reflect on the achievements of the 1956 workshop and subsequent developments in the field. Great advances in the field of AI emerged, but also moments of stalemate and backtracking. J. MOOR, *The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years*, *AI Magazine*, 27(4), 2006, 88. Available at: <https://doi.org/10.1609/aimag.v27i4.1911>.

research proceeded rapidly, carried out mainly by those who attended the Conference and their students and colleagues.

1.3. Between “AI springs” and “AI winters”

At the conference, Allen Newell and Herbert Simon presented the Logic Theorist, a programme that displayed reasoning and problem-solving capabilities⁴¹. One year later, they presented the General Problem Solver (GPS), a programme designed to solve algebraic and geometric problems, imitating human problem-solving processes. At IBM, Rochester and his colleagues developed some important AI programmes. Among them, Herbert Gelernter engineered the Geometry Theorem Prover capable of proving quite complex theorems⁴² and Arthur Samuel wrote a series of programmes for the game of checkers⁴³. After Dartmouth, McCarthy moved to MIT. In 1958 he defined the LISP language, a high-level language that became the most important language in AI programming, and he described the Advice Taker, a hypothetical programme which could be seen as the first complete AI system⁴⁴. Later, McCarthy founded the Stanford Artificial Intelligence Laboratory (SAIL).

These results led to a wave of optimism among researchers in the field, but not only. Governments also contributed to the successes of AI by investing large sums in projects connected to AI⁴⁵. Research proceeded following different approaches. For the purposes of this Thesis, two main approaches (or paradigms or perspectives) can be roughly identified: Symbolic AI and Sub-Symbolic AI (also referred to as “connectionism”)⁴⁶. Symbolic AI was mainly focused on knowledge *representation*, Sub-symbolic AI on knowledge *elaboration*⁴⁷.

⁴¹ After the Dartmouth workshop, the Logic Theorist was able to prove most of the theorems contained in the second chapter of Russel and Whitehead’s *Principia Mathematica*. See RUSSELL, NORVIG, *Artificial Intelligence. A Modern Approach*, cited, 23.

⁴² H. GELERTNER, *Realization of a geometry-theorem proving machine* in *Proceedings of the International Conference on Information Processing*, 1959, 273-282.

⁴³ Arthur Samuel is considered one of the pioneers in the field of AI. He is credited with some of the best work in the field of machine learning, using the game of draughts and chess as a vehicle for ‘training’ programmes. See J. MCCARTHY, E. FEIGENBAUM, *In memoriam*, in *AI Magazine*, 11(3), 1990, 10. See also A.L. SAMUEL, *Some Studies in Machine Learning Using the Game of Checkers*, in *IBM Journal*, 3(3), 1959, 210-219.

⁴⁴ J. MCCARTHY, *Programs with common sense*, Computer Science Department Stanford University, 1959. Available at: <http://jmc.stanford.edu/articles/mcc59/mcc59.pdf>.

⁴⁵ From the 1960s onwards, the governmental agency DARPA (an acronym for Defence Advanced Research Projects Agency) granted major funding to four of the largest American research centres: Carnegie Mellon University, MIT, Stanford University and the Stanford Research Institute. Similarly, in Great Britain the government funded.

⁴⁶ Allen Newell distinguishes between symbolic systems and continuous systems. See A. NEWELL, *Intellectual Issues in the History of Artificial Intelligence*, in F. MACHLUP, U. MANSFIELD (Eds.), *The Study of Information: Interdisciplinary Messages*, John Wiley & Sons, 1983, 187–294. Kush R. Varshney recalls that the former was focused on problem solving, psychology, performance, and serial architectures, and the latter on pattern recognition, neuroscience, learning, and parallel architectures. K.R. VARSHNEY, *Trustworthy Machine Learning*, Independently Published, 2022, 1.

⁴⁷ SARTOR, *L’informatica giuridica e le tecnologie dell’informazione*, cited, 294.

The pioneers in the field of Symbolic AI were Newell and Simon, who formulated the hypothesis that any system (human or artificial) that demonstrates intelligence must function through the manipulation of symbols⁴⁸. If the human mental activity is nothing more than a manipulation of symbols, it is thus possible to translate human cognitive processes into an explicit form, using symbols, identifying the rules of manipulation and implementing them within a machine by means of appropriate algorithms. To this end, among the different techniques, methods for automatic reasoning inspired by logic became the most relevant. Since it is the programmer who elaborates a structured knowledge representation of a specific domain, this approach is also called knowledge-based approach. The most prominent example of a concrete application of this approach is the so-called “expert system”. In the 1970s, DENDRAL, the first expert system specialised in chemical-analysis⁴⁹, and MACSYMA, a computer algebra system, were developed at MIT, and MYCIN, an expert system for medical diagnosis, was developed at Stanford University⁵⁰.

Conversely, Sub-symbolic AI is based on the idea of simulating the structure and functioning of the human brain in order to achieve intelligent behaviours. Although the brain was, and remains, a rather mysterious organ, it had long been known that signals from incoming stimuli were transmitted and altered by passing through complex networks of neurons, returning specific outputs⁵¹. What neuroscience discovered about the structure and functioning of the human brain was precisely what Frank Rosenblatt tried to imitate presenting the Perceptron in 1958, the first example of an artificial neural network⁵². The Rosenblatt’s Perceptron marked a significant milestone in AI research, being the first model to demonstrate machine learning.

⁴⁸ According to the Physical Symbol System Hypothesis «A physical symbol system has the necessary and sufficient means for general intelligent action». They specify that «By “necessary” we mean that any system that exhibits general intelligence will prove upon analysis to be a physical symbol system. By “sufficient” we mean that any physical symbol system of sufficient size can be organized further to exhibit general intelligence. By “general intelligent action” we wish to indicate the same scope of intelligence as we see in human action: that in any real situation behavior appropriate to the ends of the system and adaptive to the demands of the environment can occur, within some limits of speed and complexity». A. NEWELL, H. A. SIMON, *Computer Science as Empirical Inquiry: Symbols and Search*, *Communications of the ACM*, 19(3), 1976. Available at: <https://doi.org/10.1145/360018.360022>. By a Physical Symbolic System they meant a «machine, such as a digital computer, that is capable of manipulating symbolic data». See N.J. NILSSON, *Artificial Intelligence: A New Synthesis*, Morgan Kaufmann Publishers, 24.

⁴⁹ R.K. LINDSAY, B. G. BUCHANAN, E.A. FEIGENBAUM, J. LEDERBERG, *DENDRAL: a case study of the first expert system for scientific hypothesis formation*, *Artificial Intelligence*, 61, 1993, 209-261. See also E.A. FEIGENBAUM, B. BUCHANAN, J. LEDERBERG, *On generality and problem solving: a case study using the DENDRAL program*. Available at: <https://stacks.stanford.edu/file/druid:zr438xj0026/zr438xj0026.pdf>.

⁵⁰ See E.H. SHORTLIFFE, B.G. BUCHANAN, *A model of inexact reasoning in medicine*, in *Mathematical Biosciences*, 23, 1975, 351-379. See also B.G. BUCHANAN, E. H. SHORTLIFFE, *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, 1994.

⁵¹ G. MAIRA, *Intelligenza umana e intelligenza artificiale*, in *federalismi.it.*, 7, 2021, V–XVIII. Available at: <https://www.federalismi.it/nv14/editoriale.cfm?eid=588>.

⁵² F. ROSENBLATT, *The perceptron: A probabilistic model for information storage and organization in the brain*, in *Psychological Review*, Vol. 65, No. 6, 1958, 386-408.

Among other developments in the 1960-1970s, Prolog, a simple and intuitive logic programming language, was invented and some relevant advances also occurred in the field of natural language processing (NLP), in an attempt to enable a computer to communicate with humans in natural language. One of the greatest achievements was ELIZA, a programme developed by Joseph Weizenbaum⁵³, which represents the first chatbot in history⁵⁴.

Despite the optimism of those years, however, the goals the researchers had set themselves a few years earlier seemed unattainable. The causes of the failures lay mainly in the limitations of the technology of the time: its limited computing power and its inability to handle the amount of data required by AI systems to function properly. Moreover, the complexity of real-life problems showed that the results obtained up to that time were inadequate to solve them. At that point, not only the critics of AI⁵⁵, but also the experts in the field themselves recognised that they had underestimated the difficulties of their projects and a period of disappointment began, giving rise to the so-called “first winter of AI”.

However, as Pamela McCorduck recalls “*all sciences move in rhythms, thriving, then laying fallow (or simply assimilating what has gone before), then growing once again*”⁵⁶. Indeed, in the early 1980s, new funding arrived⁵⁷ and new advances began, both in the field of expert systems⁵⁸ and in the field of neural networks⁵⁹. Interestingly, after a period of stagnation in the research of neural networks⁶⁰, the reinvention of the backpropagation learning algorithm⁶¹, its application to problems and the

⁵³ J. WEIZENBAUM, *ELIZA—a computer program for the study of natural language communication between man and machine*, *Commun. ACM*, 9(1), 1966, 36–45.

⁵⁴ For a brief discussion on ELIZA see G. D’ACQUISTO, *Intelligenza Artificiale. Elementi*, Giappichelli, 2021.

⁵⁵ J. LIDTHILL, *Artificial Intelligence: A General Survey* in *Artificial Intelligence: a paper symposium*, Science Research Council, 1973. Available at: http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/contents.htm.

⁵⁶ MCCORDUCK, *Machines who think*, cited, 418.

⁵⁷ The Japanese government embarked on an audacious ten-year project, the Fifth-Generation project, to build intelligent computers capable of communicating with humans in natural language, understanding speech and images, learning, making decisions and carrying out any other behaviour that until then had been considered the prerogative of human reason. For a description of the Project E. FEIGENBAUM, H. SHROBE, *The Japanese National Fifth Generation Project: Introduction, survey, and evaluation* in *Future Generation Computer Systems*, 9(2), 1993, 105–117. Available at: <https://stacks.stanford.edu/file/druid:kv359wz9060/kv359wz9060.pdf>. The rest of the world, particularly Europe and the United States, reacted with an unprecedented amount of fundings and in 1983 DARPA announced the Strategic Computing Initiative.

⁵⁸ After DENDRAL, MACSYMA and MYCIN, developed in the late 1960s and early 1970s, in the 1980s R1 was the first expert system to be commercially successful. See J. MCDERMOTT, *R1: a rule-based configurer of computer systems* in *Artificial Intelligence*, 19(1), 1982, 39–88.

⁵⁹ In 1982 John Hopfield presented a model of neural network known as Hopfield Network. J.J. HOPFIELD, *Neural networks and physical systems with emergent collective computational abilities* in *Proceedings of the National Academy of Sciences of the USA*, 79(8), 1982, 2554–2558. The progresses in the field of neural network were collected in the work of James McClelland and David Rumelhart. D.E. RUMELHART, J. L. MCCLELLAND (Eds.), *Parallel Distributed Processing, Explorations in the Microstructure of Cognition I and II*, MIT Press, 1986.

⁶⁰ In 1969 M. Minsky and S. Papert published a book titled *Perceptrons: an Introduction to Computational Geometry* where they examined the capabilities of the perceptron model, highlighting its limitations in solving certain types of problems. This critique led to a widespread abandonment of neural network research that lasted well into the 1980s.

⁶¹ For a technical description of the backpropagation algorithm see D.E. RUMELHART, G.E. HINTON, R.J. WILLIAMS, *Learning representations by back-propagating errors*, in *Nature*, 323(6088), 1986, 533–536.

dissemination of the results in the collection titled *Parallel Distributed Processing*⁶², generated great excitement, establishing the central role of neural networks in AI research.

Once again, however, the wave of optimism quickly turned into a new period of reduced investments⁶³ and stagnation in AI development. For this reason, the period between late 1980s and early 1990s is commonly referred to as the “second winter of AI”. Therefore, AI research had to find new, more solid theoretical foundations and combine results from diverse disciplines in order to return to the centre stage. This, along with the availability of increasingly greater computational capabilities, more complex algorithms and larger, more affordable storage space allowed AI to flourish again and embark on what would become a new spring for AI, which continues uninterrupted to these days.

1.4. The present of AI

As Russell and Norvig point out, recent years have seen a revolution in both the *content* and the *methodology* of AI research⁶⁴. In relation to the former, current research on AI is based on existing theories, without trying to find new ones. Regarding methodology, AI research adopted the scientific method, meaning that statements are based on rigorous theorems and hard experimental evidence. Moreover, from the late 1980s, probability, decision theory and the Bayesian network formalism started playing a pivotal role in AI research leading to impressive results⁶⁵. This has become evident especially in certain AI sub-domains, such as speech recognition, machine translation and computer vision.

Starting from the late 1990s, AI gained a central role in various environments, especially in the Internet domains. Today, there seems to be no sectors in which AI does not play a prominent role. The current hype on AI can be easily explained. Computing power has increased rapidly, especially since the 2010s, and algorithms have become more complex, achieving optimal performance on several narrowly defined tasks. Moreover, at the beginning of the 21st century the society experienced a data explosion with the emergence of the World Wide Web (WWW) and various changes in the digital world, such as the emergence of cloud computing, the advent of the Internet of Things (IoT) and the

⁶² D.E. RUMELHART, J.L. MCCLELLAND, *Parallel Distributed Processing. Explorations in the Microstructure of Cognition: Foundations*, PDP Research Group, MIT Press, 1987.

⁶³ The DARPA discontinued the Strategic Computing Initiative research programme, while in Japan the Fifth-Generation project failed to achieve many of its original goals. See A. POLLACK, *Fifth Generation Became Japan's Lost Generation*, New York Times, 5 June 1992. Available at: <https://www.nytimes.com/1992/06/05/business/fifth-generation-became-japan-s-lost-generation.html>.

⁶⁴ RUSSELL, NORVIG, *Artificial Intelligence. A modern approach*, cited, 25-26.

⁶⁵ J. PEARL, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann Publishers Inc., 1988. See also P. CHEESEMAN, *In Defense of Probability*, in *Proceedings of the 9th international joint conference on Artificial intelligence - Volume 2 (IJCAI'85)*, Morgan Kaufmann Publishers Inc., 1985.

spread of social networks, have led to a continuous increase in the amount of data available at an unprecedented speed. We have now entered what is generally referred to as the “Big Data Era”⁶⁶.

The term Big Data refers to quantities of data that are very large in size and complexity. Their characteristics are generally summarised in the “3V”, namely Volume, Velocity and Variety⁶⁷: *volume* is connected to data dimension and refers to the enormous amount of data that is generated every day in various daily life activities⁶⁸; *velocity* refers to the data production rate, meaning that data is generated ever more rapidly; *variety* refers to the different types of data available today, derived from an increasing number of heterogeneous sources (structured data, semi-structured data, unstructured data, etc.). Big Data thus require non-conventional tools to extract, manage and process information within a reasonable amount of time. Recently, another “V” has been identified. *Value* concerns the potential contained in the data, which can be extracted by analysing the data and identifying trends and correlations between data.

Data constitutes the central resource for today’s AI, determining a shift from knowledge-based approaches to data-driven approaches⁶⁹. The development of an AI model relying on machine learning generally necessitates the use of three different datasets: a dataset used for the training phase of the model (*training set*); a dataset used to validate the results obtained in the training phase (*validation set*); finally, another dataset (*test set*) is used to test the model on new data that it is not familiar with. In the last few years, AI has witnessed an even greater expansion due to the amount of data available. For instance, “generative AI” or “foundational models” are trained on large data sets at scale⁷⁰. In particular, in November 2022, OpenAI launched ChatGPT, a chatbot which uses natural language

⁶⁶ One of the first contributions that highlighted both the potential, and the risks of Big Data see D. BOLLIER, *The Promise and Peril of Big Data*, Washington, DC, 2010. C. COMELLA, *Origine dei Big Data* in *Gnosis*, 2017, 130. Private enterprises also defined Big Data. See D. LANEY, *3D Data Management: controlling data Volume, Velocity and Variety*, META Group Report, File 949, 2001; M.A. BEYER, D. LANEY, *The importance of Big data: a Definition*, Gartner Analysis Report, 2012. Among others, see J. GANTZ, D. REINSEL, *Extracting value from chaos*, IDC Report, 2011; K.N. CUKIER, V. MAYER-SCHOENBERGER, *The rise of Big Data. How It's Changing the Way We Think About the World*, in *Foreign Affairs*, 92(3), 2013, 28-40; MIN CHEN, SHIWEN MAO, YUNHAO LIU, *Big data: A survey*, in *Mobile networks and applications*, 19, 2014, 171-209; OECD, *Big data: Bringing competition policy to the digital era*, 2016. M. Delmastro, A. NICITA, *Big data. Come stanno cambiando il nostro mondo*, il Mulino, 2019.

⁶⁷ Some authors envisions more characteristics, adding also veracity and value. See L. PALAZZANI, *Tecnologie dell'informazione e intelligenza artificiale*, Edizioni Studium, 2020, 26-27.

⁶⁸ The size of data produced globally today is measured in zettabytes. COM(2020), 65 final, 4. Data are collected from the digital traces linked to our daily activities, such as the use of internet-connected devices, credit card payments, digital connection with the public administration, images, videos and recordings sent and posted on social networks, medical and biometric information collected by smartwatches.

⁶⁹ «The algorithm without data is blind. Data without algorithms is dumb». CNIL’s statement reported in Future of Privacy Forum, *The Privacy Expert’s Guide to Artificial Intelligence and Machine Learning*, 2018, 8.

⁷⁰ Among others, see R. BOMMASANI ET AL., *On the opportunities and risks of foundation models*, Center for Research on Foundation Models (CRFM), 2021. Available at: <https://arxiv.org/abs/2108.07258>. M. CHEN ET AL., *Evaluating large language models trained on code*, 2021. Available at: <https://arxiv.org/abs/2107.03374>.

processing and is based on foundational models (large language models, LLMs) denoting the disruptive results that can be obtained from the use of large datasets.

The advent of Big Data has certainly favoured AI and has shown how the availability of large datasets allows AI to achieve higher performance than humans in many domains. The following picture shows that in 2023 the performance of AI models crossed the human performance (baseline set to zero) in various domains, such as reading comprehension, image recognition and language understanding.

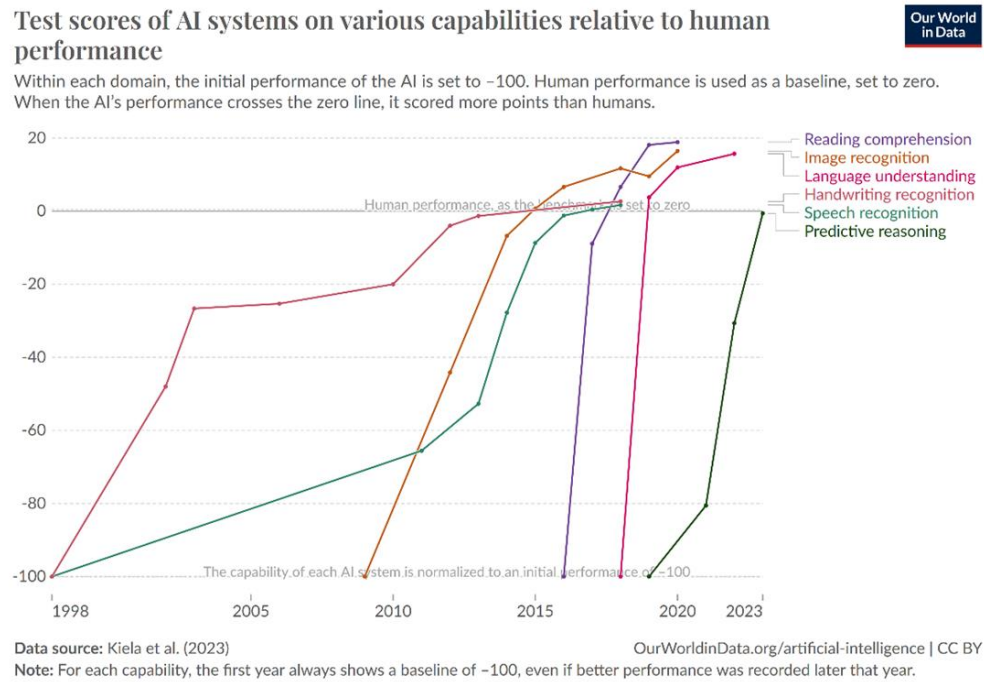


Figure 1: AI performance against human performance. Source: Our World in Data⁷¹

While the availability of data is an essential element for modern AI systems, it is the *governance* of this data that is quite complex: collecting data, integrating and preparing datasets to develop accurate AI solutions remains the main challenge for the near future⁷². As academics point out, having a large amount of data is not enough to develop successful AI. Indeed, in order to extract value from data, it is necessary to identify and extract only the data that meets certain, precise quality standards⁷³.

⁷¹ D. KIELA, T. THRUSH, K. ETHAYARAJH, A. SINGH, *Plotting Progress in AI* in *Contextual AI Blog*, 2023. Available at: <https://contextual.ai/blog/plotting-progress>.

⁷² I will discuss this aspect in Chapter 4.

⁷³ L. CAI, Y. ZHU, *The Challenges of Data Quality and Data Quality Assessment in the Big Data Era*, in *Data Science Journal*, 14, 2015, Article 2, 1-10. G. D'ACQUISTO, *Qualità dei dati e Intelligenza Artificiale: intelligenza dai dati e intelligenza dei dati*, in PIZZETTI (Ed.), *Intelligenza Artificiale, protezione dei dati personali e regolazione*, Giappichelli, 2018, 266.

1.5. The future of AI

AI systems developed in recent years show incredible improvements compared to the past decades. However, there are still several limitations that need to be overcome for future AI developments. First of all, AI systems currently being developed rely on intense computing power and in huge data center infrastructures. Energy consumption is increasing due to the huge power utilization for training and operating AI models, as well as the amount of greenhouse gases (e.g. carbon dioxide CO₂) emissions; water consumption is also increasing dramatically due to its need to cool data centers; the physical AI infrastructures (transistors, lithium batteries, GPUs) lead to both an increase in the extraction of rare earth materials and an increase in electronic waste (e-waste) due to obsolete and discarded hardware, with both environmental and healthcare risks. AI application has direct and indirect negative environmental impacts, but these topics are still the subject of little research. More research is needed to advance rigorous and empirical understandings of AI's costs for the environment and to promote a more sustainable AI⁷⁴.

Moreover, AI currently being developed still lacks good robustness. Even the smallest change in the input provided to the machine, can result in a significant reduction in its performance. An example is image recognition: if the machine is shown two images, one of which contains a disturbing element (such as a dirt stain), the two images might appear identical to the human eye, as the “extra” element would not be sufficient to alter the human perception of what the image represents. Instead, the

⁷⁴ On the topic of AI and sustainability, among others, see R. VERDECCHIA, J. SALLOU, L. CRUZ, *A systematic review of Green AI*, in *WIREs Data Mining and Knowledge Discovery*, 13(4), 2023, Article e1507, 1-43. Available at: <https://doi.org/10.1002/widm.1507>. A. DE VRIES, *The growing energy footprint of artificial intelligence in Joule*, 7(10), 2023, 2191–2194. N. BASHIR, P. DONTI, J. CUFF, S. SROKA, M. ILIC, V. SZE, C. DELIMITROU, E. OLIVETTI, *The Climate and Sustainability Implications of Generative AI*. Available at: <https://mit-genai.pubpub.org/pub/8ulgrckc/release/2>. D. PATTERSON, J. GONZALEZ, U. HOLZLE, Q. LE, C. LIANG, L.-M. MUNGUIA, D. ROTHCHILD, D.R. SO, M. TEXIER, J. DEAN, *The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink in Computer*, 55(7), 2022, 18-28. Available at: <https://doi.org/10.1109/MC.2022.3148714>. G. ZUCCON, H. SCELLS, S. ZHUANG, *Beyond CO2 emissions: The overlooked impact of water consumption of information retrieval models*, in *Proceedings of the 2023 ACM SIGIR International Conference on Theory of Information Retrieval*, 2023, 283–289. Available at: <https://doi.org/10.1145/3578337.360512>. See also the interesting work of K. CRAWFORD, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press, 2021. B. MARCHETTI, *I costi ambientali nascosti dell'intelligenza artificiale*, in *Diario di diritto pubblico*, 5 May 2024. Available at: <https://www.diariodidirittopubblico.it/i-costi-ambientali-nascosti-dellintelligenza-artificiale/>. In the USA, some members of the Congress introduced the Artificial Intelligence Environmental Impacts Act of 2024 which aims at promoting transparency, accountability, advance empirical knowledge on the AI's impact on the environment. Available at: <https://www.congress.gov/bill/118th-congress/senate-bill/3732>. I.R. HODGKINSON, N. JENNINGS, T. JACKSON, *Everyone must understand the environmental costs of AI*, October 2024. Available at: <https://oecd.ai/en/work/understand-environmental-costs>. OECD, *Measuring the environmental impacts of artificial intelligence compute and applications*, No. 341, 2022. On the environmental costs of AI see *Artificial Intelligence e Diritto - Focus on* in *BioLaw Journal*, 1, 2025, with the following contributions: B. MARCHETTI, *I costi ambientali dell'IA*, 525-527; N. RANGONE, *Intelligenza artificiale, tutela dell'ambiente e regolazione europea*, 529-548; L. PARONA, *La ponderazione dei costi ambientali nell'approvvigionamento di sistemi di intelligenza artificiale da parte delle amministrazioni pubbliche*, 549-569; L. DE GAETANO, *Il principio Do Not Significant Harm (DNSH) e i costi ambientali dell'intelligenza artificiale*, 571-590; M. MERLER, *I costi ambientali dell'intelligenza artificiale: il dibattito negli Stati Uniti*, 591-614.

machine could produce a different output for the two images, consequently providing an incorrect response⁷⁵.

Finally, existing AI systems belong to the so-called Narrow AI or Weak AI⁷⁶. Narrow AI refers to the capability of AI systems to demonstrate specific intelligence, meaning they are designed to perform well at a specific task but are unable to generalize from one problem to another. Weak AI encompasses all those practical applications where the machine must perform a specific task, such as converting speech to text, recognising faces or malignant tumours, predicting the right price to apply or recommending the next movie. The opposite of Narrow AI is Strong AI, which identifies AI systems that demonstrate general intelligence useful for solving a wide range of different tasks. When speaking of Strong AI, reference is often made to Artificial General Intelligence (AGI), which refers to a machine with a level of intelligence comparable to that of a human being. Some even speculate on a further development of AGI into what is called Artificial Superintelligence (ASI), which would entail reaching a level of general intelligence that exceeds that of humans. However, while Narrow AI is already well-developed and there are many examples of it in everyday life, Strong AI still belongs to the realm of speculation⁷⁷.

In any case, beyond extreme successes such as the creation of intelligence equal to or greater than human intelligence, even minor successes can significantly impact individuals' daily lives. These developments certainly have positive implications: AI systems are being used to ensure efficient public transportation, prevent catastrophic events, combat criminal organizations and improve healthcare. However, the risks and challenges they pose to fundamental rights cannot be underestimated.

1.6. Different ideas of AI

From an etymological point of view, *intelligence* comes from the Latin substantive *intelligentia*, a derivative of the verb *intelligere*, which means to understand, to read between the lines, to discern, and

⁷⁵ The example was provided at the *Artificial Intelligence and Ethics* meeting organised by the Collegio Clesio of the University of Trento held online on Tuesday 9 November 2021.

⁷⁶ P.N. BOUCHER, *Artificial intelligence: How does it work, why does it matter, and what can we do about it?*, Study, Think Thank European Parliament, 2020, 15-16.

⁷⁷ BOUCHER, *Artificial intelligence: How does it work, why does it matter, and what can we do about it?*, cited, 13. L. FLORIDI, *Etica dell'intelligenza artificiale*, Cortina, 2021, 218. N. BOSTROM, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, 2014. R. CELLAN-JONES, *Stephen Hawking warns artificial intelligence could end mankind* in *BBC Technology*, 2 December 2014. Available at: <https://www.bbc.com/news/technology-30290540>. Gian Maria Campedelli suggests that some recent breakthroughs have revamped enthusiasm in the research on AGI. An example is the victory of AlphaGo against the Go champion Lee Sedol in 2016, a complex game that requires more demanding abilities than those associated with chess. Another example is the release in 2022 of ChatGPT. Albeit they cannot be considered as example of AGI, they can be regarded as examples of a more comprehensive, far-reaching machine intelligence. See CAMPEDELLI, *Machine learning for criminology and crime research*, cited, 30-32.

describes the ability to establish correlations and distinctions between elements. As for the adjective *artificial*, it comes from the Latin *artificialis*, derivative of the substantive *artificium*, which is “artifice”, that which is obtained by art, as opposed to that which is by nature. AI therefore implies a non-human, artficed ability to read inside the reality.

The combination of the adjective “artificial” and the substantive “intelligence” can be contested on two main grounds. First and foremost, the reference to “intelligence” typically implies capabilities that are inherent to human beings. Consequently, applying this term to artificial entities seems to equate two fundamentally different realms – the natural and the artificial – which share minimal commonality⁷⁸. Among others, Amedeo Santosuosso argues that labelling technological artifacts as intelligent, in their current development phase, represents an “abuse”, just a “game of labels”⁷⁹. Second, intelligence is not a coherent and delineated factor, a single thing that individuals possess or do not possess, but rather a mosaic of elements that find expression in a vast array of abilities, behaviours, thoughts and emotions⁸⁰. Since defining a subjective and abstract concept like intelligence is inherently challenging, eventually resulting in inaccuracies and incompleteness⁸¹, it is not surprising that researchers continue to express significant reservations about the appropriateness of this juxtaposition. Anyhow, as Pamela McCorduck reminds, regardless of the objections and dilemmas that may legitimately flourish, “*artificial intelligence is the phrase that stuck*”⁸².

As discussed above, AI is not a single thing, it is a collective term⁸³ which encompasses a vast number of methods and techniques that have in common the goal of building systems capable of displaying capabilities that are generally associated with the concept of intelligence, such as perception,

⁷⁸ In L. PALAZZANI, *Information Technologies and Artificial Intelligence. Sfide etiche al diritto*, Edizioni Studium, 2020, 53 the author reports that a different term, such as *artificial computational system*, would have been more appropriate. Russel and Norvig also believe that a term such as *computational rationality* would have been more precise and less threatening. See RUSSEL, NORVIG, *Artificial Intelligence. A Modern Approach*, cited, 23.

⁷⁹ In A. SANTOSUOSSO, *Forum AI and Law*, in *BioLaw Journal*, 1, 2020, 469 the author argues that «what is called AI is nothing more than the broad, powerful, sometimes sophisticated development of sectoral computational capabilities. Nothing to do with the entities to which we normally attach the attribute of intelligence in our everyday lives, since before AI». Translation from Italian.

⁸⁰ Marvin Minsky in his critical review of Artificial Intelligence states: «Should we ask what intelligence “really is”? My own view is that this is more of an esthetic [aesthetic] question, or one of sense of dignity, than a technical matter! To me “intelligence” seems to denote little more than the complex of performances which we happen to respect, but do not understand». M. MINSKY, *Steps toward Artificial Intelligence* in *Proceedings of the IRE*, 49(1), 1961, 8-30. Available at: <https://ieeexplore.ieee.org/document/4066245>. For further discussion of the absence of unanimity of the concept of intelligence see R.L. GREGORY, *Intelligence, The Oxford Companion to the mind*, Oxford University Press, 1987, 375-379.

⁸¹ See Joint Research Centre, *Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*, Luxembourg, 2020, 7. P. DE BOECK, *Intelligence, Where to Look, Where to Go?*, in *Journal of Intelligence*, 1(1), 2013, 5-24.

⁸² «In my view it is a wonderfully appropriate name, connoting a link between art and science that as a field AI indeed represents». MCCORDUCK, *Machines who think*, cited, 115.

⁸³ See A. D’ALOIA, *Il diritto verso “il mondo nuovo”. Le sfide dell’Intelligenza Artificiale* in A. D’ALOIA (Ed.), *Intelligenza artificiale e diritto. How to regulate a new world*, Franco Angeli, 2020, 19. See also BOUCHER, *Artificial intelligence: How does it work, why does it matter, and what can we do about it?*, cited, 1.

CHAPTER 1

programming, reasoning and learning⁸⁴. As highlighted above, the developments that led AI systems to display to a greater or lesser extent these capabilities have been the result of a combination of insights from the most diverse disciplines: from philosophy to mathematics, from economy to neuroscience, from psychology to computer engineering, from cybernetics to linguistics. The multidisciplinary nature of AI research has led to the development of different ideas about AI among experts, each of whom adopts a different perspective on what can be considered a demonstration of intelligence by electronic calculators.

In their textbook *Artificial Intelligence. A modern approach*⁸⁵, Russell and Norvig attempt to make sense of this complex plethora of perspectives by imagining diverse categories within which AI research can be divided. To do so, they cross two distinctions: the first one, between “thinking” and “acting”, distinguishing between studies that aims at creating intelligent entities based on modes of thinking and reasoning and those that focus on behaviour; the second one, between “humanity” and “rationality” as the metric for measuring performance. Based on these differences it is thus possible to identify four different approaches: 1) systems that think like human beings, 2) systems that think rationally, 3) systems that act like human beings and 4) systems that act rationally.

	Human performance	Rational performance
<i>Thought processes</i>	Thinking humanly (1)	Thinking rationally (2)
<i>Action/behaviour</i>	Acting humanly (3)	Acting rationally (4)

Figure 2: Different approaches in AI research based on S. Russell, P. Norvig, *Artificial Intelligence. A modern approach*, cited.

The first category refers to the branch of AI research grounded in the aspiration to model and replicate human modes of thinking in artificial systems⁸⁶. This line of research has its basis in cognitive sciences, an interdisciplinary field in which various theories on the functioning of the human mind have been developed and used as models to be implemented in computer programmes. An example of this approach is represented by the research of Newell and Simon, whose aim with the General Problem Solver was not merely to have their program solve problems correctly, but also to compare the path of

⁸⁴ J. MCCARTHY, *What is artificial intelligence?*, Stanford, 2007, <http://jmc.stanford.edu/articles/whatisai.html>. «*Artificial intelligence is the study of machines that exhibit traits associated with a human mind such as perception, learning, reasoning, planning, and problem solving*». VARSHNEY, *Trustworthy Machine learning*, cited, 1.

⁸⁵ RUSSELL, NORVIG, *Artificial Intelligence. A Modern Approach*, cited.

⁸⁶ J. HAUGELAND, *Artificial Intelligence: the very idea*, MIT Press, 1985. R.E. BELLMAN, *An introduction to Artificial Intelligence: Can computer think?*, Boyd & Fraser Publishing Company, 1978.

its reasoning process to traces of human subjects solving the same problems. John Haugeland's research can be linked to this approach⁸⁷.

The thinking rationally approach, which has its basis in logic, encompasses the strand of research that has worked on the formalisation of reasoning mechanisms into general, abstract rules and the development of notation systems, derived from mathematical logic, with which concepts and relations between them can be described and codified within a machine⁸⁸.

The third category of AI research is aimed at replicating human behaviour⁸⁹. This approach has its roots in Turing's paper *Computing Machinery and Intelligence* published in 1950⁹⁰. He proposed an operational definition of AI by presenting a test, subsequently called the Turing Test, that attempts to determine whether a computer can show intelligence. According to Turing, a computer can be considered intelligent if it behaves in a way that is indistinguishable from a human being⁹¹: the computer therefore passes the test if a human interrogator communicating in written form with other actors (a human being and a computer) is unable to distinguish, based on the answers given, whether the answers come from the computer or the human being⁹². In the following decades, the Turing Test spurred a strong debate and was strongly criticised by John Searle, who formulated the so-called Chinese Room Argument, a famous thought experiment designed to challenge the idea that AI – specifically, a computer running the right program – could *truly* understand language or possess a mind⁹³. Although the Turing Test is still relevant today, AI researchers have devoted little effort to passing it, focusing instead on different approaches to construct intelligent artefacts.

The last category includes AI research focused on creating agents that act in a way that achieves the best outcome, using rationality as the metric of performance. Russell and Norvig favour this last approach, expressly stating that an AI system is first and foremost a “rational agent”⁹⁴. Specifically, an AI system is described as an agent that acts within its environment in order to achieve specific goals

⁸⁷ The author refers to AI as «*the existing new effort to make computers think (...) machines with minds, in the full and literal sense*».

⁸⁸ E. CHARNIAK, D. MCDERMOTT, *Introduction to artificial intelligence*, Addison-Wesley, 1985; P.H. WINSTON, *Artificial Intelligence*, Addison-Wesley, 1992. The authors describe their field of study specifically as «the study of mental faculties through the use of computational models».

⁸⁹ R. KURZWEIL, *The Age of Spiritual Machines*, Orion, 1999. Raymond Kurzweil described AI as «the art of creating machines that perform functions that require intelligence when performed by people».

⁹⁰ A.M. TURING, *Computing Machinery and Intelligence*, in *Mind*, 49, 1950, 433-460.

⁹¹ RUSSELL, NORVIG, *Artificial Intelligence. A Modern Approach*, cited, 2.

⁹² For a computer to pass the Turing Test is far from simple, requiring the computer to possess different capabilities: natural language processing, in order to communicate effectively; knowledge representation, to store what it knows; automated reasoning, to use its knowledge to answer the question; machine learning, to adapt to new circumstances. Although Turing acknowledged that the machines of his time were unable to pass the test, he was optimistic that by the end of the 20th century computers would become sophisticated enough to be indistinguishable from humans in conversation.

⁹³ J. SEARLE, *Minds, Brains, and Programs*, in *Behavioral and Brain Sciences*, 3(3), 1980, 417-424.

⁹⁴ For a more detailed discussion of the concept of rationality in AI systems see RUSSELL, NORVIG, *Artificial Intelligence. A Modern Approach*, cited, 36-39.

and an agent can be said to be rational if, given its knowledge, it always does “the right thing”, *i.e.* the one that makes it achieve the highest degree of success⁹⁵. When the environments in which AI systems operate are complex, the concept of “bounded rationality” introduced by Herbert Simon in the 1950s becomes relevant: an agent can make rational, albeit not fully optimal decisions, considering the limitation of information available, cognitive limitations⁹⁶ and time constraints⁹⁷. The performance of the rational agent is then assessed by the desirability of the sequences of actions generated by the agent when operating in its environment⁹⁸.

The conception of AI based on rational behaviour (the “acting rationally” approach) is currently prevalent in the scientific literature for several reasons. First, it allows for a wider variety of approaches to be included in the concept of AI, considering that logical inference may not be the only path an agent has to follow in order to successfully negotiate its environment. Second, it anchors the notion of “intelligence” to a less abstract concept, which is the concept of “rationality”, meaning the ability to choose the best action to take to achieve a certain goal, given the available resources. The “rational agent” approach is the basis of various documents that have attempted to define AI for normative purposes.

2. Technical coordinates on artificial intelligence

In order to fully understand the definition of an AI system encompassed in the AI Act, it is necessary to provide some basic technical notions⁹⁹. It must be noted that the following discussion is inevitably a simplification of the complex AI technical landscape, but it is particularly useful for non-AI experts looking at this technology from a legal perspective.

First and foremost, the difference between “AI system” and “AI model” is relevant. From a legal perspective it is important to distinguish between the two, since the AI Act establishes binding rules mainly at the level of the AI system. However, in some cases, *i.e.* when general-purpose AI models

⁹⁵ «For each possible percept sequence, a rational agent should select an action that is expected to maximize its performance measure, given the evidence provided by the percept sequence and whatever built-in knowledge the agent has». RUSSELL, NORVIG, *Artificial Intelligence. A Modern Approach*, cited, 37.

⁹⁶ In relation to AI systems, cognitive limitation can be better understood as limited computing power.

⁹⁷ *Bounded Rationality* in *Stanford Encyclopedia of Philosophy*, 30 November 2018. Available at: <https://plato.stanford.edu/entries/bounded-rationality/>. A. TSAOUSSI, *Bounded Rationality* in J. BACKHAUS (Ed.) *Encyclopedia of Law and Economics*, 2014. Available at: https://doi.org/10.1007/978-1-4614-7883-6_106-1. The concept of bounded rationality represents the basis of heuristics.

⁹⁸ RUSSELL, NORVIG, *Artificial Intelligence. A Modern Approach*, cited, 37.

⁹⁹ For the following discussion reference is made to P. TRAVERSO, *Breve introduzione tecnica all'Intelligenza Artificiale*, in *DPCE online*, 1, 2022, 155-167. OECD, *Framework for the Classification of AI systems*, *OECD Digital Economy Papers*, No. 323, OECD Publishing, Paris, 2022. Available at: <https://doi.org/10.1787/cb6d9eca-en>. RUSSELL, NORVIG, *Artificial Intelligence. A Modern Approach*, cited.

come into play, it also introduces binding rules at the level of the AI model¹⁰⁰. An “AI system” is a broader concept than an “AI model”, the latter being solely a component of an AI system. More precisely, an AI system includes an AI model along with additional components that allow it to function effectively in real-world applications, such as sensors, interfaces, traditional software, and to perform its tasks¹⁰¹. Consequently, while AI models represent essential components of AI systems, they do not in themselves constitute complete AI systems, as they invariably require additional components to operationalise their functionality and interface with virtual or physical environments. Nevertheless, to accurately classify an AI system, it is necessary to consider the AI model around which the system is built¹⁰². An “AI model” represents the core of an AI system and it can be described as a specific mathematical representation that maps inputs with relevant outcomes¹⁰³. More precisely, an AI model is an actionable representation of the environment of an AI system which describes the environment’s structures and dynamics¹⁰⁴.

The concept of “algorithm” is also relevant¹⁰⁵. An algorithm is a well-defined sequence of elementary actions that an executor should perform in order to arrive, within a limited period of time, at the solution to a computational problem¹⁰⁶. Hence, an algorithm is the procedure whose correct execution makes it possible to go from certain inputs to certain outputs. Therefore, while an algorithm is a sequence of instructions (dynamic process) that builds and optimizes the model, the AI model is the resulting entity (static entity) built and optimized by the algorithm. The latter is then capable of making predictions or analysing new data. An AI system is thus said to infer outputs when it takes input from

¹⁰⁰ AI Act, Chapter V: Articles from 51 to 56.

¹⁰¹ The vast majority of AI systems that are deployed to operate in real-world contexts are composite systems. See D. FERNÁNDEZ LLORCA ET AL., *An interdisciplinary account of the terminological choices by EU policymakers ahead of the final agreement on the AI Act: AI system, general purpose AI system, foundation model, and generative AI*, 7.

¹⁰² OECD, *Framework for the Classification of AI systems*, cited, 42.

¹⁰³ AI model is a «physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, process or data» (ISO/IEC 22989:2022, Information Technology — Artificial Intelligence — Concepts and Terminology, 3.1.23) and a machine learning model is a «mathematical construct that generates an inference or prediction based on input data» (ISO/IEC 22989, cited, 3.3.7).

¹⁰⁴ The original text is: «A model is an actionable representation of all or part of the external environment of an AI system that describes the environment’s structure and/or dynamics». OECD, *Scoping the OECD AI principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)*, *OECD Digital Economy Papers*, No. 291, OECD Publishing, Paris, 2019. Available at: <https://doi.org/10.1787/d62f618a-en>.

¹⁰⁵ There is no perfect coincidence between the definitions of algorithm. See SARTOR, *L’informatica giuridica e le tecnologie dell’informazione*, cited, 96-97. See D. BALDINI, D., M. DE BENEDETTO, *The open texture of ‘algorithm’ in legal language*, in *AI & Soc*, 40, 2025, 1643-1654.

¹⁰⁶ The idea of defining precise methods in order to elaborate information has a long history. Euclide between the IV and III century b.C. invented a famous algorithm to calculate the greatest common divisor. The concept *algorithm* derives from the Persian mathematician Muhammad ibn Mūsā al-Khwārizmī, lived in the IX century a.C., whose works contributed to the introduction of decimals and algebra in Europe. The technical definition of algorithm is «1. finite set of well-defined rules for the solution of a problem in a finite number of steps 2. sequence of operations for performing a specific task 3. finite ordered set of well-defined rules for the solution of a problem». ISO/IEC/IEEE 24765:2017, *Systems and software engineering — Vocabulary*, 3.124. Available at: <https://www.iso.org/obp/ui/en/#iso:std:iso-iec-ieee:24765:ed-2:v1:en>.

the environment and computes a result by processing this input through one or more models and algorithms¹⁰⁷.

Furthermore, there is a close relationship between algorithms and software that deserves to be further investigated. Software can be defined as “*a set of computer programs, procedures, and possibly associated documentation and data*”¹⁰⁸. A software presents different components¹⁰⁹: a conceptual component, a textual component, a material component and a virtual component. Without going into the details of every single part, the first two are the most relevant for the purpose of this analysis. The conceptual part consists of algorithms and software architectures, which are then expressed by computer programmes. Computer programmes, which represents the textual component with the relevant documentation, are instructions expressed in a programming language¹¹⁰ that can be executed automatically by a computer¹¹¹. The instructions written by the programmer that guide the execution of a computer programme are the source code, while the object code is the automatic translation of the source code into machine-readable language. The communicative function of software is of utmost importance for legal purposes¹¹².

The reference to the “environment” is of utmost importance in relation to AI systems¹¹³. The environment – a real or virtual, observable or partially observable space¹¹⁴ – is both perceived and influenced by the AI system: an AI system detects data from the environment via sensors (e.g. microphones, cameras, etc.), analyses it and finally instructs actuators to perform operations based on its processing in the environment¹¹⁵. For example, an AI system used for security purposes can identify visual data through cameras (sensor), process it and reach conclusions (e.g. attempted intrusion), based

¹⁰⁷ OECD, *Updates to the OECD's definition of an AI system explained*, November 2023. From a legal point of view, the academic world has questioned the legal status of an «algorithm», in particular its classification as an «administrative act». The discussion is polarised, as it is possible to identify two opposing theses. In relation to those who consider an algorithm an administrative act see U. FANTIGROSSI, *Automazione e pubblica amministrazione. Profili giuridici*, il Mulino, 1993, 51-62; A. CONTALDO, L. MAROTTA, *L'informatizzazione dell'atto amministrativo: cenni sulle problematiche in campo*, in *Il Diritto dell'informazione e dell'informatica*, 3, 2002, 571-601. For the opposite view cfr. A.G. OROFINO, *La patologia dell'atto amministrativo elettronico: sindacato giurisdizionale e strumenti di tutela*, in *Foro Amministrativo-C.d.S.*, 9, 2022, 2256-2281.

¹⁰⁸ ISO/IEC 5055:2021, Information technology — Software measurement — Software quality measurement — Automated source code quality measures, 4.13. Available at: <https://www.iso.org/obp/ui/en/#iso:std:iso-iec:5055:ed-1:v1:en>.

¹⁰⁹ For the following discussion reference is made to SARTOR, *L'informatica giuridica e le tecnologie dell'informazione*, cited, 114.

¹¹⁰ There are numerous programming languages. The most common are Java, Python, JavaScript, R, C/C++.

¹¹¹ SARTOR, *L'informatica giuridica e le tecnologie dell'informazione*, cited, 114.

¹¹² Access to the source code is considered one of the conditions for the admissibility of the use of AI in the Italian public administration, which is necessary to ensure compliance with the principle of knowability and comprehensibility. I deal with this topic in Chapter 4, Part I.

¹¹³ For technical discussion on the nature of environment see RUSSELL, NORVIG, *Artificial Intelligence. A Modern Approach*, cited, 40-46.

¹¹⁴ OECD, *Scoping the OECD AI Principles*, cited, 7.

¹¹⁵ TRAVERSO, *Breve introduzione tecnica all'Intelligenza Artificiale*, cited, 158; See SARTOR, *L'informatica giuridica e le tecnologie dell'informazione*, cit., 65; RUSSELL, NORVIG, *Artificial Intelligence*, cited, 34: «An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators».

on which it can trigger alarms (actuator). The following picture shows a detailed conceptual view of an AI system.

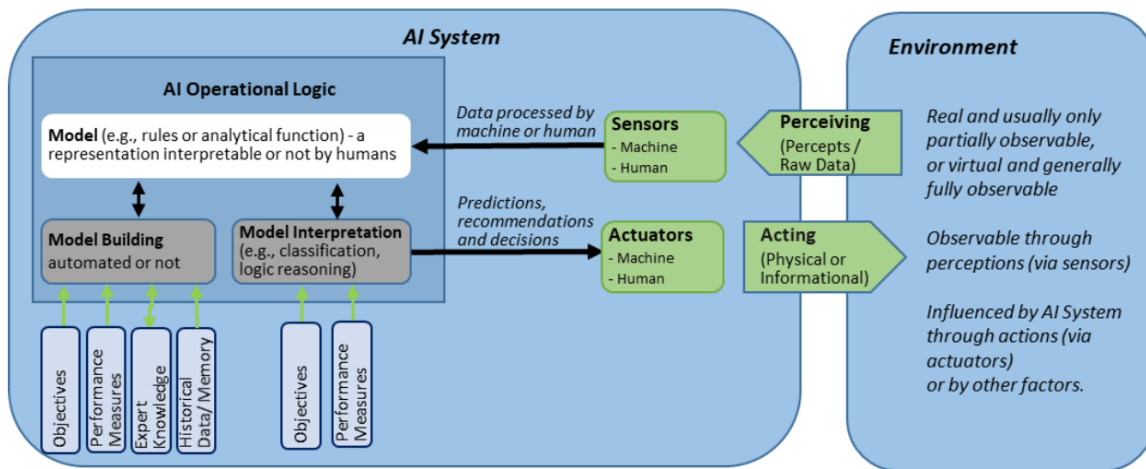


Figure 3: Conceptual view of an AI system. Source: OECD (2019)

As the figure shows it is important to distinguish between two different phases: the “build phase”, which represents the pre-development phase, and the “model interpretation” which is the use phase of the model after its deployment.

2.1. Model building

First and foremost, it is necessary to focus on the model-building process. The ways in which an AI model is constructed can be roughly grouped into two categories: human-encoded knowledge and machine learning. As Giovanni Sartor recalls, the last decades have witnessed a central shift in the development of AI systems that moves from logic to statistics as a way of processing information and from knowledge representation to machine learning as a way of generating models¹¹⁶.

2.1.1. Knowledge representation

Early research into the development of intelligent entities employed techniques to generate models emulating high-level functions of the brain, notably formal knowledge representation, logic and inference. Since there are currently no computer systems that can understand the semantics of the natural language, the operations a system can do depend on the words (= sequence of alphanumeric

¹¹⁶ SARTOR, *L'informatica giuridica e le tecnologie dell'informazione*, cited, 299.

characters) and the syntactic structures linking them. These systems work based on form and not content, thus the reasoning is said to be “formal”.

This type of approach is described as “symbolic” because it makes use of symbolic reasoning: it is therefore sufficient to have assumptions, recognise the linguistic structure (form) of the assumptions and apply a rule of inference to the elements contained in that structure. This approach has its roots in Aristotle’s syllogism which provided models of argumentative structures that always produced correct conclusions when given correct premises¹¹⁷. Symbolic approach requires experts to encode their knowledge in a precise set of rules, meaning to explicitly describe the variables and their interrelationships, which a computer can follow step-by-step to arrive at a solution¹¹⁸. The most basic AI models are therefore a series of if-then-else statements with rules explicitly programmed by a human expert.

The significant advantage of symbolic approaches lies in their intelligibility by human beings: the path followed by the machine to transform inputs into outputs is clear and understandable to programmers¹¹⁹. This makes it possible to thoroughly check the system’s operations, identify errors and potentially intervene in the process. However, such systems have objective limitations: first, the presence of an expert is always required to instruct them and improve their operation by means of continuous interventions; second, these systems apply their knowledge base, thus if this does not include all the information necessary to solve a problem, the output would be incorrect¹²⁰; third, symbolic AI systems can become computationally expensive as the number of symbols and rules increases¹²¹. Nevertheless, symbolic AI systems still show their usefulness in areas where the rules are more rigid and less ambiguous. These systems are thus referred to as Good-old-fashioned-AI (GOFAD)¹²².

2.1.2. Machine learning

As the codification of knowledge or symbolic representation of the task to be solved proved particularly laborious, a new approach began to emerge which emphasises the importance of learning

¹¹⁷ The typical example is: «Socrates is a man» and «all men are mortal», «therefore, Socrates is mortal». Generalising, X is Y, z is a X, therefore z is Y.

¹¹⁸ See BOUCHER, *Artificial intelligence: How does it work, why does it matter, and what can we do about it?*, cited, 2.

¹¹⁹ In relation to symbolic AI some authors said: «Neither machines nor programs are black boxes; they are artefacts that have been designed, both hardware and software, and we can open them up and look inside». A. NEWELL, E. SIMON, *Computer science as empirical inquiry: Symbols and search*, in *Communications of the ACM*, 19(3), 114.

¹²⁰ As Russell and Norvig point out «It is not easy to take informal knowledge and state it in the formal terms required by logical notation, particularly when the knowledge is less than 100% certain». RUSSELL, NORVIG, *Artificial Intelligence. A Modern Approach*, cited, 4.

¹²¹ Cfr. BOUCHER, *Artificial intelligence: How does it work, why does it matter, and what can we do about it?*, cited, 1.

¹²² Cfr. HAUGELAND, *Artificial Intelligence. The Very Idea*, cited.

and adaptation, rather than relying solely on predefined rules. The term machine learning (ML) was coined by Arthur Samuel in the late 1950s¹²³ and it was further elaborated by Tom Mitchell in 1997¹²⁴. ML represents a data-driven approach that allows AI systems to learn and improve from data without being explicitly programmed¹²⁵. Hence, in ML-based systems knowledge is not encoded by experts, but is constructed directly by the system on the basis of the data provided. More precisely, ML algorithms analyse the data, create a specific mathematical function that links certain inputs to outputs in order to perform the given task and generalise it so that it can be applied to solve similar tasks based on new, unseen data¹²⁶. In ML-based systems it is thus necessary to distinguish the *learning algorithm* that constructs the model and the *model* itself that generates outputs from new inputs¹²⁷.

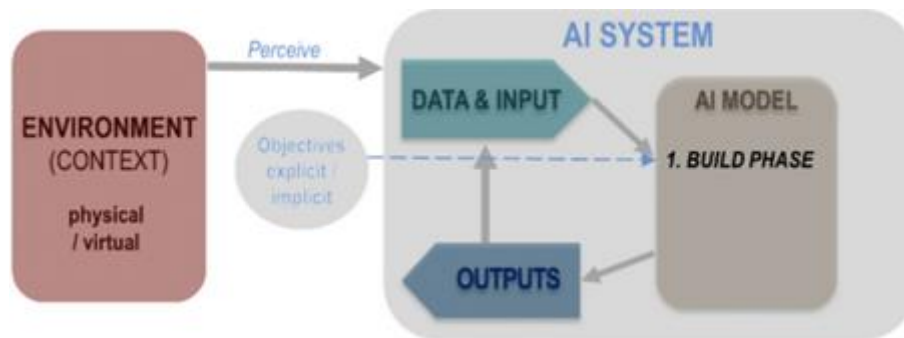


Figure 4: Representation of the build phase (pre-deployment) of machine learning. Source: OECD (2024)

Some clarifications are necessary¹²⁸. First, there is no single approach to learning to be used in ML. For the purpose of this Thesis, three main categories can be roughly identified: supervised learning, unsupervised learning and reinforcement learning¹²⁹. In supervised learning algorithms, labelled data

¹²³ A.L. SAMUEL, *Some Studies in Machine Learning Using the Game of Checkers*, in *IBM Journal of Research and Development*, 3(3), 1959, 210-229. Available at: <https://ieeexplore.ieee.org/abstract/document/5392560>. Samuel defined ML as the «field of study that gives computers the ability to learn without being explicitly programmed».

¹²⁴ T.M. MITCHELL, *Machine Learning*, McGraw-Hill, 1997. «A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E».

¹²⁵ This definition is the result of Samuel and Mitchell's definitions of ML. «Machine Learning is the study of computer algorithms that improve automatically through experience». This definition is available at <https://www.ieeesmc.org/technical-activities/cybernetics/machine-learning/>.

¹²⁶ SARTOR, *L'informatica giuridica e le tecnologie dell'informazione*, cited, 311-315; D'AQUISTO, *Intelligenza Artificiale. Elementi*, cited, 127.

¹²⁷ OECD, *Explanatory memorandum on the updated OECD definition of an AI system*, *OECD Artificial Intelligence Papers*, No. 8, OECD Publishing, Paris, 2024. Available at: <https://doi.org/10.1787/623da898-en>.

¹²⁸ For the following discussion reference is made to G. D'AQUISTO, *Intelligenza Artificiale. Elementi*, 127-183.

¹²⁹ D'AQUISTO, *Intelligenza Artificiale. Elementi*, cited, 127-129. I.H. SARKER, *Machine Learning: Algorithms, Real-World Applications and Research Directions*, in *SN Computer Science*, 2, 2021, Article 160, 1-21. Accessible at: <https://doi.org/10.1007/s42979-021-00592-x>. For further discussion see also CAMPEDELLI, *Machine learning for criminology and crime research*, cited, 35-40.

sets are used to train the model. Labelled datasets mean that input data are matched to correct outputs or response values (input-output sample pairs). The learning algorithm attempts to create a model to map inputs to outputs and the model is then used to predict the response to any new set of input data. A relevant example of supervised learning is an image classification AI system, whereby each image is matched with the correct label. Unsupervised learning algorithms receive large amounts of unlabelled data, *i.e.* without labelled responses, and find implicit and hidden patterns in the data, without requiring prior labelling by a human expert. AI systems for anomaly detection generally use unsupervised learning. Reinforcement learning algorithms allow the machine to maximise its performance in each environment through positive or negative feedback, *i.e.* rewards or penalties. Second, there is no single technique and algorithm to be used in ML¹³⁰. Among the methods, classification and regression techniques are used in supervised learning; clustering is the most common technique, albeit not the only one, in unsupervised learning. Classification is a special form of learning in which the agent classifies incoming data into categories based on the characteristics (variables) of the data. An example of classification tasks is anomaly detection, where the objective is to classify data as normal or anomalous. Multiple algorithms can be used for classification tasks (Support Vector Machine, Decision Tree, K-Nearest Neighbors, etc.). Regression algorithms aim to identify a relationship between a dependent variable (target variable) and one or more independent variables (predictor variables). The objective is to determine the most suitable function to describe the relationship between these variables. Crime pattern analysis and prediction generally use regression models¹³¹. Clustering is a machine learning technique in which the agent observes data and tries to group them into clusters with similar attributes. Clustering algorithms are extremely useful for finding hidden patterns and associations in datasets. New tasks can then be performed based on the results obtained.

2.1.3. Deep learning

The current frontier of AI is deep learning. Deep learning is a branch of ML that is based on neural network architectures. It has its roots in the studies of the mid-twentieth century conducted by

¹³⁰ For the following discussion reference is made at MatLab. Available at: https://se.mathworks.com/discovery/machine-learning.html?s_tid=srchtitle_site_search_4_machine%20learning.

¹³¹ S. DEGADWALA, D. VYAS, A. R. RAVAL, M. SONI, *Crime Pattern Analysis and Prediction Using Regression Models*, in *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, Erode, India, 2023, 771-776. DOI: 10.1109/ICSSAS57918.2023.10331747.

McCulloch and Pitts (1943)¹³² and by Rosenblatt (1958)¹³³ who attempted to imitate the structure of the human brain and its functioning to create machine intelligence. A neural network consists of interconnected nodes (parameters), called “artificial neurons”, in a layered structure that relate inputs to outputs. At least three different layers can be identified, namely the input layer, the output layer and the hidden layer in between. The term “deep” usually refers to the number of hidden layers of the neural network, which can be hundreds or even thousands¹³⁴ and are linked by mathematical functions.

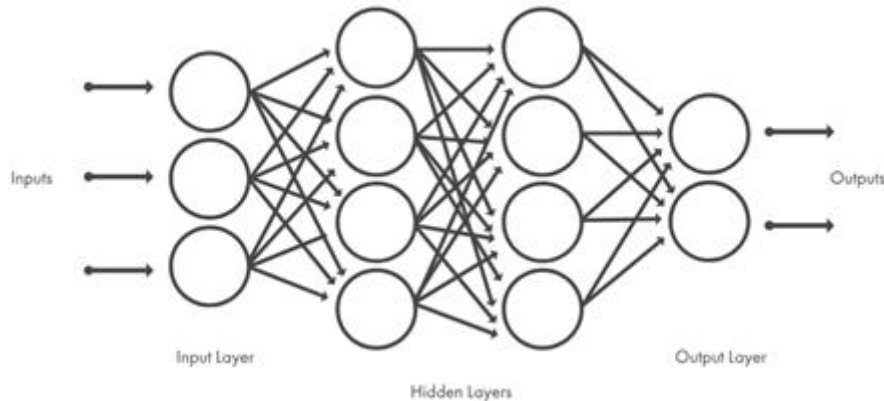


Figure 5: Typical neural network architecture. Source: MATLAB

There are different types of neural networks. The most common are convolutional neural networks (CNNs)¹³⁵ and recurrent neural networks (RNNs)¹³⁶. Convolutional neural networks, for example, are extremely accurate for image recognition tasks, as demonstrated by AlexNet in ImageNet Large Scale Visual Recognition Challenge in 2012¹³⁷.

Deep learning is excellent in predictive tasks, as it can map variables (inputs) to outcomes with even very complex relationships between them. However, achieving these results comes at a cost: deep learning relies on a vast number of layers and parameters that demand large amounts of training data. This complexity makes it extremely difficult for humans to trace how a neural network arrives at a

¹³² MCCULLOCH, PITTS, *A logical calculus of ideas immanent in nervous activity*, cited.

¹³³ ROSENBLATT, *The perceptron*, cited.

¹³⁴ MathWorks, *What Is Deep Learning?*. Available at: https://se.mathworks.com/discovery/deep-learning.html?s_tid=srchtitle_site_search_3_deeplearning. P. TRAVERSO, *Breve introduzione tecnica all'Intelligenza Artificiale*, 161-163. For technical discussion see I. GOODFELLOW, Y. BENGIO, A. COURVILLE, *Deep Learning*, Boston, 2016.

¹³⁵ An interesting explanation of CNNs can be found in M. MITCHELL, *L'intelligenza artificiale. Una guida per esseri umani pensanti*, Einaudi, 2022.

¹³⁶ For a first technical discussion R.M. SCHMIDT, *Recurrent Neural Networks (RNNs): A gentle Introduction and Overview*. Available on: arXiv:1912.05911.

¹³⁷ A. KRIZHEVSKY, I. SUTSKEVER, G.E. HINTON, *ImageNet Classification with Deep Convolutional Neural Networks*, in *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 2012, 1097-1105.

given output from its inputs, a limitation that has earned deep learning the label of “Black Box” due to its lack of explainability¹³⁸.

In conclusion, leaving the technical details to the extensive literature on the topic¹³⁹, what is relevant for my research is that the selection of approaches (knowledge representation, machine learning, deep learning), approaches to learning (supervised, unsupervised, reinforcement learning), techniques (classification, regression, clustering) and algorithms (Decision Trees, Neural Networks, Linear Regression, etc.) to train a model depends on various aspects: what is the problem to be solved, what size and type of data are to be used, what insights are to be gained from the data and how these insights will be used. Depending on the above considerations, the model itself may be more or less accurate or interpretable by the programmer, giving rise to various legal, ethical and social problems.

2.2. Inference phase

In AI research, inference can be described as the process to move from certain known premises to conclusions, where a premise is either a fact, a rule, a model, a feature or raw data¹⁴⁰. The concept of inference generally refers to the phase after the deployment of the AI model, when the model, whether encoded by human experts or learned from data, is used to obtain outputs (e.g. predictions, recommendations, decisions etc.) in relation to new, unseen data¹⁴¹. In this second phase, the model is then put into action in the reference environment on real data to produce usable results.

¹³⁸ Y. BATHAEE, *The Artificial Intelligence Black Box And The Failure Of Intent And Causation*, in *Harvard Journal of Law & Technology*, 31(2), 2018, 889-938.

¹³⁹ I.H. WITTEN, E. FRANK, *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann, 2005. J. HAN, J. PEI, M. KAMBER, *Data mining: concepts and techniques*, Elsevier, 2011. Interesting insights can be found in MATLAB. Available at: https://se.mathworks.com/solutions/artificial-intelligence.html?s_tid=hp_hero_ai. For further discussion on different algorithms in machine learning, Carnegie Mellon University's Auton Lab maintains a set of tutorials. Available at: <https://autonlab.org/tutorials.html>. RUSSELL, NORVIG, *Artificial Intelligence. A modern approach*, cited; D'AQUISTO, *Intelligenza artificiale. Elementi*, cited.

¹⁴⁰ OECD, *Framework for the Classification of AI systems*, cited, 48. ISO standards defines inference as the «process of reasoning by which conclusions are derived from known premises». ISO 22989:2022, Information technology — Artificial intelligence — Artificial intelligence concepts and terminology, clause 3.1.7. The OECD described inference as the «process by which humans and/or automated tools derive an outcome from the model». OECD, *Scoping the OECD AI Principles*, cited, 7-8.

¹⁴¹ OECD, *Framework for the Classification of AI systems*, cited, 48.

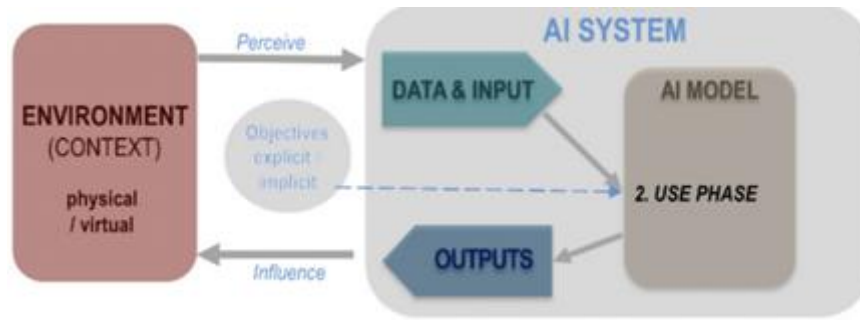


Figure 6: Representation of the use phase (post-deployment) of machine learning. Source: OECD (2024)

It should be noted that while AI models may remain unchanged after deployment once the build phase is concluded, some AI models based on ML methods can continue to adapt after the initial build phase, improving their performance by interacting directly with new input data¹⁴². Starting from the same model (trained and validated), different inference strategies can be used to derive different results (e.g. Bayesian inference, causal inference). A fundamental difference is the one between deterministic inference and probabilistic inference (statistical inference)¹⁴³. The main difference between the two is that the latter involve randomness to better handle incomplete information, uncertainty, unpredictability and data variation, which are typical of real-life situations, while the former do not involve it. Depending on the type of inference, one can distinguish between deterministic and probabilistic models (stochastic models¹⁴⁴). While deterministic models follow precise rules and generate a single result from the same input, in the case of stochastic models, given the same initial conditions and parameters, the results may be different¹⁴⁵. The choice between the two depends on the environment in which the AI model operates. The difference between deterministic and stochastic models is relevant in legal considerations because some key properties of AI models, such as the degree of explainability, robustness, testing and testability depend on the differences discussed above.

¹⁴² Tom Mitchell's definition of machine learning is enlightening on this point. Central to this definition is the concept of "experience", which means that the aim is to create programmes that learn from experience to improve the results of their actions, *i.e.* to improve their performance in solving a given task.

¹⁴³ OECD, *Framework for the Classification of AI systems*, cited, 48-49.

¹⁴⁴ RUSSELL, NORVIG, *Artificial Intelligence. A modern approach*, cited, 43 the use of the term "stochastic" generally implies that uncertainty about outcomes is quantified in terms of probabilities. For further discussion on the probabilistic reasoning, see D'AQUISTO, *Intelligenza artificiale. Elementi*, cited, 51.

¹⁴⁵ OECD, *Framework for the Classification of AI systems*, cited, 48-49. Many machine learning algorithms leverage stochasticity to enhance their learning processes and improve performance (optimization).

3. From causation to correlation

The increase in the amount of available data and the general shift towards ML- or DL-techniques have led to a different paradigm or epistemological approach to make sense of the world¹⁴⁶: *correlation* instead of *causation*¹⁴⁷. In all sciences, causation indicates a direct cause-and-effect relationship between two variables: it means that changes in one variable result in a change of the other. Conversely, correlation refers to a statistical association between two variables. More precisely, correlation is a measure, expressed as a number, that describes the *size* and *direction* (positive or negative) of a relationship between two or more variables¹⁴⁸. It indicates that when one variable changes, the other *tends* to change as well. However, it does not automatically mean that variation in one of the variables is the cause of variation in the values of the other variable. More simply, correlation does not imply causation¹⁴⁹.

Correlation generally represents the basis of AI systems that generate predictions¹⁵⁰. In relation to ML, Giovanni Sartor defines correlation as the result of probabilistic relationships between input data and possible outcomes: if certain inputs result in a higher probability of a certain outcome, the relationship is considered a *positive* correlation; conversely, if an input results in a lower probability of a certain outcome, it is referred to as a *negative* correlation¹⁵¹. These correlations are incorporated in the model. Exemplifying the above¹⁵², in ML-based AI systems, the learning algorithm builds a predictive model which is based on correlations between certain characteristics of the cases considered (e.g. attributes of the defendants) and the answers inferred based on these characteristics (e.g. criminal behaviour). Once the AI model is trained, the predictive model can be used in relation to new cases and can evolve over time, improving its performance. The system tends to generate a positive prediction¹⁵³ when input

¹⁴⁶ On this topic, see R. KITCHIN, *Big Data, New Epistemologies and Paradigm Shifts*, in *Big Data & Society*, Vol. 1, No. 1, 2014. <https://doi.org/10.1177/2053951714528481>; D. BOYD, K. CRAWFORD, *Critical questions for big data*, in *Information, Communication and Society*, 15(5), 2012, 662–679. V. MAYER-SCHNBERGER, K. CUKIER, *Big Data: A Revolution That Will Transform How We Live, Work and Think*, Houghton Mifflin Harcourt, 2013.

¹⁴⁷ “Big Data refers to novel ways in which organizations, including government and businesses, combine diverse digital datasets and then use statistics and other data mining techniques to extract from them both hidden information and surprising correlations”. I. S. RUBENSTEIN, *Big Data: The End of Privacy or a New Beginning?*, in *International Data Privacy Law*, 3(2), 2013, 74. M. L. AMBROSE, *Lessons from the Avalanche of Numbers: Big Data in Historical Perspective*, in *I/S: A Journal Of Law And Policy For The Information Society*, 11(2), 2015, 201.

¹⁴⁸ Australian Bureau of Statistics, Correlation and causation. Available at: <https://www.abs.gov.au/statistics/understanding-statistics/statistical-terms-and-concepts/correlation-and-causation>

¹⁴⁹ For a detailed discussion on the concept of causation and correlation see G.U. YULE, *On the Methods Of Measuring Association Between Two Attributes*, in *Journal Of The Royal Statistical Society*, 75(6), 1912, 579–652.

¹⁵⁰ Systems based on ML have been referred to as «prediction machines». See A. AGRAWAL, J. GANS, A. GOLDFARB, *Prediction machines. The Simple Economics of Artificial Intelligence*, Harvard Business Review Press, 2018.

¹⁵¹ SARTOR, *L’informatica giuridica e le tecnologie dell’informazione*, cited, 315–316.

¹⁵² Reference is made at F. LAGIOIA, G. SARTOR, *Il sistema COMPAS: algoritmi, previsioni, iniquità* in *XXVI Lezioni di Intelligenza Artificiale. Saggi a margine del ciclo seminariale “Intelligenza Artificiale e diritto”*, Giappichelli, 2020, 227.

¹⁵³ In this context positive is not synonym of favourable.

data is positively correlated by the learning algorithm with a certain outcome and a negative prediction when input data and the outcome are negatively correlated.

In the criminal justice system, for example, predictive software may be used to estimate the risk of reoffending. The software may generate a positive prediction (*i.e.* another crime will be committed) if the defendant has a history of drug offenses and the model learned that there is a positive correlation between such histories and future criminal behaviour. Conversely, a negative prediction may be generated if the defendant has no such prior offenses in his criminal history. In general, all relevant correlations that the system must consider are aggregated into a score expressing the probability that, in the specific case, the prediction is positive instead of negative.

For the purpose of this Thesis, the term *prediction* is used accordingly to Sartor's works¹⁵⁴. With this concept Sartor generally describes the outcomes of ML-systems. More precisely, the author states that predictions, when ML is involved, have a meaning of their own, different from the meaning that is attributed to that term in everyday language. In fact, ML-predictions do not always concern the future (e.g. predicting a defendant's future behaviour) but may also concern the present or the past. For example, a software used for object recognition it is said to predict the content of the image with a certain probability.

The difference between causation and correlation is of utmost importance from a legal perspective. Indeed, although the estimation of correlations in the data has yielded interesting insights leading to improve decision-making in many sectors without regard for causal mechanism¹⁵⁵, failure to understand the root causes of outcomes can lead to erroneous or even prejudicial decisions.

An enlightening example is represented by risk-prediction tools in the U.S. healthcare system to target patients for high-risk care management programs. The underlying objective of these tools is that individuals with greater care needs would benefit from the programme more than others. As stated in research conducted on the topic¹⁵⁶, identifying those patients is a *causal* inference problem, given the fact that the sicker an individual is, the highest should be the algorithm score and thus the greatest should be the benefit from the programme. However, the authors found out that the algorithm used health-care costs as a proxy for illness. They thus highlighted that health-care costs are not always related to the individual's degree of illness. Indeed, in the case of subjects qualified as "Black", higher rates of illness did not lead to higher costs because of reasons unrelated to health-care needs, namely

¹⁵⁴ SARTOR, *L'informatica giuridica e le tecnologie dell'informazione*, cited; LAGIOIA, SARTOR, *Il sistema COMPAS: algoritmi, previsioni, iniquità*, cited.

¹⁵⁵ For example, predicting infection in premature infants in intensive care 24 hours before overt symptoms appear. See M. BLOUNT, M.R. EBLING, J.M. EKLUND ET AL., *Real-time analysis for intensive care: development and deployment of the Artemis analytic system*, in *IEEE Engineering in Medicine and Biology Magazine*, 29(2), 2010, 110-118.

¹⁵⁶ The reference goes to Z. OBERMEYER, B. POWERS, C. VOGELI, S. MULLAINATHAN, *Dissecting racial bias in an algorithm used to manage the health of populations*, in *Science*, 366, 2019, 447-453.

barriers to health-care access, inadequate health care and lack of insurance, which are generally connected to race and socio-economic status. Therefore, the result was a biased outcome produced by the algorithm for some patients (the Black ones), because, with the same algorithm score, the subjects qualified as “Black” had a significantly worse health condition than the subjects qualified as “White”. In the case of Black patient there was correlation but not causation¹⁵⁷.

The results of this research are not only relevant for health care decisions. Rather, they are also relevant to the use of risk-prediction tools in the criminal justice system or in law enforcement contexts. In all these areas, when correlations are mistakenly taken as causation, the consequences for the individual can be devastating¹⁵⁸. Indeed, being able to predict an outcome, because of estimated correlations between the data, is not the same as understanding what causes this outcome. Therefore, it is imperative that the decision-maker is aware of the difference between causation and correlation and that risk-prediction tools are generally based on the latter. In some sectors, calls for a different approach, namely causal-AI, are rising¹⁵⁹. Indeed, it has been emphasised that some use cases require causal modelling, instead of predictive modelling, aimed at finding instances of cause-and-effect from the data rather than only correlative patterns¹⁶⁰.

4. A legal definition of AI

Despite the abundance of AI definitions in textbooks, institutional documents and articles, numbering in the hundreds, none of them appears to capture the full breadth and depth of the vast world of AI. Given the inherent challenges in formulating a definition of AI that is impervious to criticism regarding its incompleteness, vagueness or excessive generalization, a universally accepted notion of AI remains

¹⁵⁷ S.K. SGAIER, V. HUANG, G. CHARLES, *The Case for Causal AI*, in *Stanford Social Innovation Review*, 18(3), 2020, 50–55. Available at: <https://doi.org/10.48558/KT81-SN73>.

¹⁵⁸ One of the most famous cases in ProPublica, J. ANGWIN, J. LARSON, S. MATTU, L. KIRCHNER, *Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks*, 23 May 2016. Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

¹⁵⁹ L. CAVIQUE, *Implications of causality in artificial intelligence*, in *Frontiers in Artificial Intelligence*, 7, 2024 Article 1339936, 1–4. Available at: <https://doi.org/10.3389/frai.2024.1439702>.

¹⁶⁰ VARSHNEY, *Trustworthy Machine Learning*, cited, 21. For further discussion see Chapter 8 of the book. I will further discuss the topic in Chapter 4.

elusive¹⁶¹. However, from a legal perspective defining concepts is crucial as it delineates the scope of regulatory policies and actions¹⁶².

Most legal documents do not provide a general definition of AI¹⁶³, rather they focus on the concept of “AI system”, which represents a more concrete and appropriate term in legal contexts¹⁶⁴. Although a legal definition is crucial for regulatory purposes, it does not necessarily capture the ontological meaning of the concept it defines, in this case AI. Instead, the aim is to focus on the specific elements that characterise the systems that legislators want to regulate, clarifying which systems fall within the scope of a regulation and are subject to the requirements and obligations set out therein¹⁶⁵.

Since the research focuses on the European landscape, it is necessary to examine the institutional definitions provided in this territorial context¹⁶⁶. The first European definition of an “AI system” is due to the European Commission in 2018¹⁶⁷. Later, in 2019, the High-Level Expert Group on Artificial Intelligence (AI HLEG), set up by the European Commission, proposed a more comprehensive, technical and detailed definition¹⁶⁸. In the same year, the Organization for Economic Cooperation and

¹⁶¹ The Joint Research Centre published an interesting report whose aim was to identify all the domains that AI encompasses and to seek an operational definition of AI by collecting and analysing many existing definitions. S. SAMOILI, M. LOPEZ COBO, E. GOMEZ GUTIERREZ, G. DE PRATO, F. MARTINEZ-PLUMED, B. DELIPETREV, *AI Watch. Defining Artificial Intelligence*, Publications Office of the European Union, Luxembourg, 2020; S. SAMOILI, M. LÓPEZ COBO, B. DELIPETREV, F. MARTÍNEZ-PLUMED, E. GÓMEZ, G. DE PRATO, *AI Watch. Defining artificial intelligence 2.0. towards an operational definition and taxonomy for the ai land scape*, Publications Office of the European Union, Luxembourg, 2021. On the definition of AI see also H. SHEIKH, C. PRINS, E. SCHRIJVERS, *Mission AI. The New System Technology*, Springer, 2023, 15-41.

¹⁶² D. MURDICK, J. DUNHAM, J. MELOT, *AI Definitions Affect Policymaking*, Center for Security and Emerging Technology, June 2020. Available at: <https://cset.georgetown.edu/research/ai-definitions-affect-policymaking/>.

¹⁶³ The ISO standard describes AI as a «<discipline> research and development of mechanisms and applications of AI systems». ISO/IEC 22989:2022, Information Technology — Artificial Intelligence — Concepts and Terminology.

¹⁶⁴ The ISO standard 22989/2022 describes an «AI system» as an «engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives».

¹⁶⁵ J. ELLUL, *Should we regulate Artificial Intelligence or some uses of software?*, in *Discover Artificial Intelligence*, 2, 2022, Article 5. Available at: <https://doi.org/10.1007/s44163-022-00021-9>. Legal definitions tend to remain under-examined by legal and social science scholarship. R. RICHARDSON, *Defining and Demystifying Automated Decision Systems*, in *Maryland Law Review*, 81(3), 2022, 790. Available at: <https://digitalcommons.law.umaryland.edu/mlr/vol81/iss3/2>. For a comprehensive analysis of the legal definition of AI see C. TRINCADO CASTÀN, *The legal concept of artificial intelligence: the debate surrounding the definition of AI System in the AI Act*, in *Biolaw Journal*, 1, 2024, 305-344. See also J. SCHUETT, *Defining the Scope of AI Regulations*, in *Law, Innovation and Technology*, Legal Priorities Project Working Paper Series, 15(1), 2023, 64. Available at: <https://www.tandfonline.com/doi/epdf/10.1080/17579961.2023.2184135?needAccess=true>.

¹⁶⁶ All the other definitions derived from the scientific literature and/or non-binding documents are not fit for purpose. C. AIKEN, *Classifying AI Systems*, Center for Security and Emerging Technology, November 2021. <https://doi.org/10.51593/20200025>.

¹⁶⁷ «Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions –with some degree of autonomy –to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications)». European Commission, *Communication: Artificial Intelligence for Europe, COM(2018) 237 final, 25 April 2018*. European Commission, *Communication: Coordinated Plan on Artificial Intelligence, COM(2018) 795 final, 7 December 2018*. The reference to «intelligent behaviours» would have been problematic in a legal document since it is connected to the concept of *intelligence* which is an extremely vague term.

¹⁶⁸ «Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition,

Development (OECD) provided a different definition of an AI system¹⁶⁹. Although the OECD definition was non-binding, it was one of the most influential worldwide, including in the European scene.

In 2021, the European Commission proposed the draft regulation on AI in which it provided a legal definition of an AI system that denotes common elements and differences with that of the OECD¹⁷⁰. Since the publication of the draft, during the trilogue, the definition has changed several times. In November 2022, the Council of the European Union adopted its General Approach (Council's Position) introducing some significant changes in the Commission's Proposal¹⁷¹. The European Parliament published its position in June 2023, in which several changes were made to the original definition, also departing from the Council's Position¹⁷². In parallel with the deliberations at the EU level, the OECD revised its initial definition. Finally, to align EU's definition with other international organisations working on AI, facilitating international convergence, ensuring legal certainty and encourage broad acceptance, the final and definitive version of the AI Act opted for a definition of AI system that is closer to the OECD's revised version:

interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions. As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber physical systems)». High Level Expert Group on Artificial Intelligence, *A Definition of AI: Main Capabilities and Disciplines*, December 2018. Available at: https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf.

¹⁶⁹ «Machine-based systems that can, for a given set of human defined objectives, make predictions, recommendations or decisions, influencing real or virtual environments, and are designed to operate with varying levels of autonomy». OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, May 2019 (revised May 2024). Available at: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.

¹⁷⁰ «AI system means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with». European Commission, *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*, COM(2021) 206 final, 2021/0106(COD), 21 April 2021. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>.

¹⁷¹ «AI system means a system that is designed to operate with elements of autonomy and that, based on machine and/ or human-provided data and inputs, infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts». Council of the European Union, Proposal for a Regulation on Artificial Intelligence, General Approach, Interinstitutional File: 2021/0106(COD), 25 November 2022. Available at: <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf>.

¹⁷² «AI system means a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments». Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html.

“An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment” (OECD, *Recommendation of the Council on Artificial Intelligence*, 2023)¹⁷³

“‘AI system’ means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments” (UE, *AI Act*, 2024)¹⁷⁴

The final definition encompassed in Article 3(1) AI Act entered into application on 2 February 2025¹⁷⁵. It aims to provide the flexibility to accommodate the rapid technological developments in the field¹⁷⁶ and stresses some important aspects of an AI system, which differentiate it from non-AI systems¹⁷⁷. To this end, Recital 12 offers some important insights. First and foremost, the AI Act adopts a definition that encompasses the two main phases described above: the pre-deployment (or “building”) phase and the post-deployment (or “use”) phase of the system.

At the end of the trilogy, an AI system is defined as a “machine-based system”, moving away from the Proposal’s concept of “software”¹⁷⁸ and specifying the general reference to “system” provided in the Council’s Position. As specified in the Commission’s *Guidelines on the definition of an Artificial*

¹⁷³ OECD, *Updates to the OECD’s definition of an AI system explained*, cited.

¹⁷⁴ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/182 (Artificial Intelligence Act). Official Journal of the European Union, L series, 1-144, 12 July 2024. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689.

¹⁷⁵ In the same day the European Commission published the Guidelines on the definition of an AI system. See European Commission, *Guidelines on the Definition of an Artificial Intelligence System Established by Regulation (EU) 2024/1689 (AI Act)*, COM(2025) 924 final, 6 February 2025.

¹⁷⁶ Additional techniques and applications will likely emerge in the future, as happened with the advent of ChatGPT in November 2022. The definition of AI must therefore be future proof in order to encompass systems which will be the result of future developments in order to prevent the immediate obsolescence of the legislation. See A. RENDA, A. ENGLER, *What’s in a name? Getting the definition of Artificial Intelligence right in the EU’s AI Act*, in *CEPS EXPLAINERS*, 2023. Available at: <https://www.ceps.eu/ceps-publications/whats-in-a-name/>.

¹⁷⁷ The importance to differentiate between AI systems and non-AI systems has been stressed by the European Commission. See European Commission, *Impact Assessment of the Regulation on Artificial Intelligence*, SWD(2021) 84 final, 2021.

¹⁷⁸ The Commission referred to «software» in order to align the definition in the AI Act with already existing references to «software» in other EU legislative acts, such as the Medical Device Regulation. See D. FERNÁNDEZ LLORCA, E. GÓMEZ, I. SÁNCHEZ, G. MAZZINI, *An interdisciplinary account of the terminological choices by EU policymakers ahead of the final agreement on the AI Act: AI system, general purpose AI system, foundation model, and generative AI in Artificial Intelligence and Law*, 33(4), 2024, 875–888. Further discussion on the difference between «software» and «AI» see TRINCADO CASTÀN, *The legal concept of artificial intelligence*, cited, 330-333.

Intelligence System (hereinafter “Guidelines”) the term “machine based” refers to the fact that AI systems are developed and run on machines which can be both software and hardware components¹⁷⁹. Furthermore, the reference to the objectives was amended: instead of referring to “human-defined objectives”, as in the Council’s Position, the final version refers to “explicit or implicit objectives”, meaning that an AI system operates according to objectives which can be explicitly defined by humans, implicitly encompassed in human-defined rules or in training data or even not fully known in advance¹⁸⁰. The objectives of an AI system are internal to the system “referring to the goals of the tasks to be performed and their results”¹⁸¹ and can differ from the “intended purpose of the AI system in a specific context”¹⁸² which “includes the context in which the system is designed to be deployed and how it must be operated”¹⁸³.

The characteristics of an AI system can be summarised in four main elements¹⁸⁴: (i) inference capability, (ii) autonomy, (iii) adaptability and iv) capability to influence the physical or virtual world. *Autonomy* (ii) is a vague concept, but it refers to the ability of the system to interact with its external environment autonomously. Recital 12 specifies that it means that AI systems should have some degree of independence of action from human involvement¹⁸⁵. The reference to different “levels of autonomy” is both positive, since it renders autonomy “a gradual and non-dichotomous variable”¹⁸⁶, and negative, since it leaves some ambiguity in the definition, requiring further guidance to specify what is the minimum level of autonomy to meet the criterion. Certainly, AI systems that are designed to operate solely with full manual human involvement and intervention are outside the scope of the Regulation¹⁸⁷. The doubt thus was whether an “expert system” (which is an example of knowledge-based AI system) falls within the definition of an AI system of the AI Act. Doubts were finally resolved by the Guidelines, which specify that “*an expert system following a delegation of process automation by*

¹⁷⁹ European Commission, *Guidelines on the Definition of an Artificial Intelligence System*, cited, 2.

¹⁸⁰ OECD, *Explanatory memorandum on the updated OECD definition of an AI system*, cited, 7.

¹⁸¹ See European Commission, *Guidelines on the Definition of an Artificial Intelligence System*, cited, 4.

¹⁸² AI Act, Recital 12.

¹⁸³ See European Commission, *Guidelines on the Definition of an Artificial Intelligence System*, cited, 4.

¹⁸⁴ In the Guidelines published by the European Commission in February 2025 seven main elements are identified: (1) a machine-based system; (2) that is designed to operate with varying levels of autonomy; (3) that may exhibit adaptiveness after deployment; (4) and that, for explicit or implicit objectives; (5) infers, from the input it receives, how to generate outputs (6) such as predictions, content, recommendations, or decisions (7) that can influence physical or virtual environments. *Ibidem*, 2.

¹⁸⁵ The test states: «*AI systems are designed to operate with varying levels of autonomy, meaning that they have some degree of independence of actions from human involvement and of capabilities to operate without human intervention*».

¹⁸⁶ TRINCADO CASTÁN, *The legal concept of artificial intelligence*, cited, 329.

¹⁸⁷ Autonomy can be described as a spectrum where at one extreme are systems which are designed to perform all tasks though manually operated functions and at the other extreme are systems that are capable to operate without any human involvement or intervention. “*For example, a system that requires manually provided inputs to generate an output by itself is a system with ‘some degree of independence of action’, because the system is designed with the capability to generate an output without this output being manually controlled, or explicitly and exactly specified by a human*”. European Commission, *Guidelines on the Definition of an Artificial Intelligence System*, cited, 3.

humans that is capable, based on input provided by a human, to produce an output on its own such as a recommendation is a system with ‘some degree of independence of action’”¹⁸⁸.

Following Recital 12, *adaptiveness* (iii) relates to specific AI systems that can continue to adapt after the initial training phase and after deployment, improving their performance by interacting directly with new inputs¹⁸⁹. The reference thus goes to AI systems that show self-learning capabilities and can improve their performance after deployment (*i.e.* machine learning).

Autonomy and adaptiveness are considered as different concepts in the AI Act, but they should be more intertwined than envisaged, the former requiring adaptability to be fully realised. Again, the AI Act considers adaptiveness as an “optional” characteristic of an AI system, rather than a fundamental one, as the modal verbs “may” in Article 3(1) and “could” in Recital 12 underline. This makes adaptability a rather weak requirement in the definition and, given the above, autonomy may also be weakened. On the contrary, in the OECD definition both autonomy and adaptability seem to be inherent in all AI systems, different levels of autonomy and adaptability being possible, but not the absence of one.

Undoubtedly, the central core of an AI system is its *capability to infer* (i). This notion encompasses a variety of inference-enabling techniques. These include machine-learning approaches, as well as deep learning and certain knowledge-based approaches, with the sole exception of traditional software systems relying solely on fixed rules and systems executing operations exclusively on the basis of instructions defined by natural persons. As stated in the Guidelines, the capability to infer is “*a key, indispensable condition that distinguishes AI systems from other types of systems*”¹⁹⁰. The concept of inference encompassed in the AI Act is broader than the one I described in paragraph 4.2. On one hand, it refers to the building phase. Indeed, Recital 12 stresses that the capability to infer refers “*to a capability of AI systems to derive models or algorithms, or both, from inputs or data*” which underlines “*the relevance of the techniques used for building a system*”¹⁹¹. On the other hand, it refers to the use phase. Indeed, Recital 12 describes the capability to infer as the “*process of obtaining the outputs, such as predictions, content, recommendations, or decisions, which can influence physical and virtual environments*” which refers to the application of the AI model to new, unseen data, enabling machines to obtain outputs. The broad meaning of the concept of inference included in the AI Act is clearly underlined by the Guidelines which state that “*The terms ‘infer how to’, used in Article 3(1) and clarified in recital 12 AI Act, is broader than, and not limited only to, a narrow understanding of the*

¹⁸⁸ See *Ibidem*.

¹⁸⁹ The test states: «*The adaptiveness that an AI system could exhibit after deployment, refers to self-learning capabilities, allowing the system to change while in use*».

¹⁹⁰ *Ibidem*, 5.

¹⁹¹ *Ibidem*.

*concept of inference as an ability of a system to derive outputs from given inputs, and thus infer the result. Accordingly, the formulation used in Article 3(1) AI Act, i.e. ‘infers, how to generate outputs’, should be understood as referring to the building phase, whereby a system derives outputs through AI techniques enabling inferencing”*¹⁹².

Despite its centrality, the reference to inferential capacity changed several times throughout the trilogue, especially with regard to AI techniques that enable inference. Originally, the Commission’s Proposal stated that an AI system could have been developed with different “*techniques and approaches listed in Annex I*”, which mentioned machine learning, logic- and knowledge-based approaches and statistical modelling¹⁹³. However, in December 2022 the Council adopted its General Approach, with which it suppressed Annex I and incorporated the techniques and approaches included therein into the definition itself through the phrase “*infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge-based approaches, and produces system-generated outputs*”. Later, the European Parliament proposed a different definition, omitting the reference to AI ability to infer how to achieve its goals¹⁹⁴. In the end, the final version is closer to the Council’s Position. The definition provided by Article 3(1) is inclusive enough to encompass different techniques that enable inference in AI systems, including both logic- and knowledge-based approaches, where the model is manually encoded, and learning approaches, where the model is trained from data, such as machine learning and deep learning. In any case, Recital 12’s list of techniques enabling inference is non-exhaustive in nature, as evidenced by the verb “include”¹⁹⁵.

Furthermore, Recital 12 explicitly states that the AI system’s ability to infer transcends basic data processing, “*enabling learning, reasoning or modelling*”¹⁹⁶. This means it must go beyond simple, fixed-rule automation and be capable of inferring outputs, rather than merely executing hard-coded instructions and it must generate predictions or decisions with a varying degree of autonomy. The reference to the ability of an AI system to learn, reason and model is crucial to distinguish it from simpler software and programming approaches. Although this reference may narrow the scope of the

¹⁹² *Ibidem*.

¹⁹³ The reference to «statistical approaches» as a third category of techniques would have been extremely problematic. First, because both machine learning and some knowledge-based systems rely on statistical approaches. Second, because almost any other kind of software relies on statistical approaches. See P. GLAUNER, *An Assessment of the AI Regulation Proposed by the European Commission*, in S. EHSANI, P. GLAUNER, P. PLUGMANN, F. M. THIERINGER (Eds.), *The Future Circle of Healthcare: AI, 3DPrinting, Longevity, Ethics, and Uncertainty Mitigation*, Springer, 2022, 119-127. With the reference to Annex I, which could have been amended by means of delegated acts, the Commission’s aim was to clarify and exemplify the definition itself.

¹⁹⁴ The European Parliament’s definition was criticised since it was not clear in distinguishing AI systems from non-AI systems.

¹⁹⁵ See T. BOBEV, *Defining AI in the AI Act: Pin the Tail on the System*, 2 April 2024. Available at: <https://www.law.kuleuven.be/citip/blog/defining-ai-in-the-ai-act-pin-the-tail-on-the-system/>.

¹⁹⁶ Many changes have been introduced in Recital 6 (current Recital 12) stressing that the definition should be based on the key characteristics of AI, which distinguish it from simpler software systems and programming approaches.

definition, it contributes to legal uncertainty as to which systems fall within the definition. The only certainty is that the definition provided by Article 3(1) AI Act “*should not cover systems that are based on the rules defined solely by natural persons to automatically execute operations*”¹⁹⁷, which therefore are outside the scope of the Regulation.

The positive aspect of the final version of the definition is that it focuses on *how* the system *works*, instead of which techniques are used for its development, in order to be classified as an AI system¹⁹⁸. Therefore, systems that may not be considered AI from a technical or ontological point of view can be considered AI systems for the scope of the AI Act. For example, given the wording of the definition, borderline cases in advanced statistical modelling could be considered AI for the purposes of the AI Act if they demonstrate the key characteristics of AI systems set out in the definition¹⁹⁹.

In conclusion, the boundaries of the definition of an AI system in the AI Act remain uncertain, given the reference to indeterminate and broadly interpretable concepts, such as autonomy, inference, reasoning. As I will discuss in Chapter 2, when concrete cases are examined, determining whether a system falls within the definition of AI is not always straightforward. In some instances, simpler AI systems are difficult to distinguish from traditional software systems – which fall outside the scope of the AI Act – or other systems that present a limited capacity to analyse patterns or autonomously adapt their outputs. In order to limit the uncertainty, the Commission with the technical assistance of the AI Office provided guidance to assist in the practical application of the provision²⁰⁰. Moreover, it is expected that the Court of Justice will intervene to promote the uniform interpretation of the definition among Member States, also on the basis of preliminary ruling proceedings²⁰¹. How these elements are to be interpreted will be a central aspect in determining when a machine-based system does or does not fall within the scope of the AI Act.

¹⁹⁷ AI Act, Recital 12.

¹⁹⁸ TRINCADO CASTÀN, *The legal concept of artificial intelligence*, cited, 332. Accordingly, the Guidelines state that “*The determination of whether a software system is an AI system should be based on the specific architecture and functionality of a given system and should take into consideration the seven elements of the definition laid down in Article 3(1) AI Act*”. European Commission, *Guidelines on the Definition of an Artificial Intelligence System*, cited, 12.

¹⁹⁹ A. GELMAN, *Bayesian statistics: What’s it all about?*, 13 December 2016. Available at: <https://statmodeling.stat.columbia.edu/2016/12/13/bayesian-statistics-whats/>.

²⁰⁰ Article 96 of the AI Act states that the Commission shall develop Guidelines on the practical implementation of the Regulation. In particular, the Commission shall develop these Guidelines in relation to “*(f) the application of the definition of an AI system as set out in Article 3, point (1)*”.

²⁰¹ TFEU, Article 267.

5. An overview of the AI lifecycle

In this last paragraph, I briefly mention some aspects that will be explored in more detail in the following Chapters in order to provide a general overview that may be useful in understanding the later discussion.

AI is an intrinsically complex technology. Therefore, when an organisation decides to develop or implement an AI system it is essential that every aspect is adequately taken into consideration. This is especially true if the organization is a public authority, as the use of AI in public administration poses many challenges from a legal point of view²⁰². Once the problem to be solved has been in principle identified, the problem owner (e.g. the public authority) must first decide whether the AI system is the most suitable solution to solve the problem itself²⁰³. Only if the AI system is considered the most appropriate one, then the problem owner can proceed further and specify both the problem and the relevant aspects of the AI system to be implemented. The sequential and iterative process that leads from the definition of the problem to the development of the AI system that aims to solve it is called the AI system lifecycle. For the purposes of this Thesis, phases involved in the AI system lifecycle can be identified as follows²⁰⁴: i) planning, ii) data understanding, iii) data preparation, iv) modelling, v) evaluation, vi) deployment and monitoring. An optional phase is the withdrawal or deactivation of the system.

²⁰² The use of AI, in particular ML, by the public sector has given rise to a vast literature. Among others, see C. COGLIANESE, *Administrative Law in the Automated State*, in *Daedalus*, 150(3), 2021, 104–120. C. COGLIANESE, D. LEHR, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, in *Georgetown Law Journal*, 105, 2017, 1147-1223.

²⁰³ Varshney warns: do not solve problems that would lead to harms for someone or some group. Strating from this statement different sub-questions arise: Does the problem have a solution? Does the imagined solution cause harm to anyone? Is the solution acceptable? In the end, is it worth solving the problem? Specifically, is machine learning the most appropriate solution to the problem? Can the problem be solved with simple programming? In the end, should we automate processes just because we can?

²⁰⁴ The reference model is the Cross-Industry Standard Process for Data Mining (CRISP-DM). C. SHEARER, *The CRISP-DM model: the new blueprint for data mining*, in *Journal of Data Warehousing*, 5(4), 2000, 13–22. The six-phase process is recalled in VARSHNEY, *Trustworthy Machine Learning*, cited, 14. Some authors identify more phases of the AI lifecycle: 1) Inception/Analysis, 2) Data acquisition and preparation, 3) Development, 4) Verification and validation, 5) Deployment, 6) Operation and monitoring, 7) Continuous validation, 8) Re-evaluation, 9) Retirement. See European Data Protection Supervisor, *Guidance for Risk Management of Artificial Intelligence systems*, 11 November 2025.

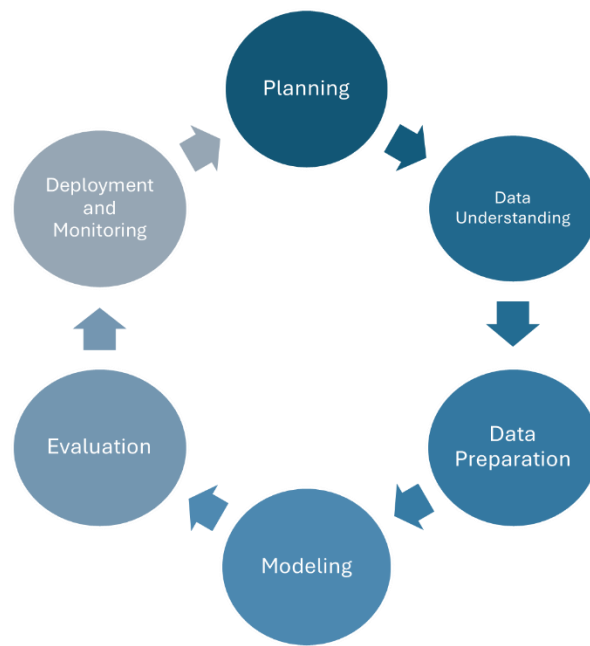


Figure 7: AI lifecycle based on Cross-Industry Standard Process for Data Mining (CRISP-DM). Own elaboration

As Kush R. Varshney recalls, developing a trustworthy AI, and ML in particular, is essential when AI is used in public sector organizations²⁰⁵. In order to achieve this goal, it is crucial to follow the process outlined above step-by-step and to actively involve different personas, although with different roles and responsibilities, in every stage of the process. Personas involved include, *inter alia*, the problem owner, who is the public authority whose problem must be solved, the data provider, the data scientist, the data engineer and other stakeholders, such as the civil society, the academic community and the labour representatives²⁰⁶.

In the planning phase, personas involved must focus on the definition and specification of the problem that will be addressed with the AI solution and articulate all relevant requirements of the AI system, such as concepts, objectives, underlying assumptions, context in which the system is intended to operate and metrics of success. The correct definition of the problem is crucial, as the type of data to be collected and the type of algorithms best suited to solve the problem depend on it. In this step, the

²⁰⁵ An enlightening analysis can be found in VARSHNEY, *Trustworthy Machine Learning*, cited, 14-22. Varshney highlights the importance of realizing trustworthy ML to build trust and use AI for social good and analyses concepts for developing accurate, fair, robust, explainable, transparent, inclusive, empowering and beneficial ML systems.

²⁰⁶ In February 2025 the Agency for Digital Italy (AgID) launched a public consultation on the Guidelines for the adoption of AI in Public Administration, which have not yet been adopted. AgID, *Guidelines for the adoption of Artificial Intelligence in Public Administration*, Resolution No. 17/2025, 18 February 2025. Available at: <https://www.agid.gov.it/en/intervention-areas/artificial-intelligence>. The Guidelines adopt as a reference model the AI lifecycle defined by the OECD, which recalls the AI lifecycle described above. See also OECD, *Scoping the OECD AI Principles*, cited, 13. For a more detailed analysis see D. DE SILVA, D. ALAHAKOON, *An artificial intelligence life cycle: From conception to production*, in *Patterns*, 3(6), 2022, 100489, 1-13. Available at: <https://www.sciencedirect.com/science/article/pii/S2666389922000745>.

problem owner, notably the public organisation, plays a central role, exercising the discretion granted by the law. However, a dialogue between all interested parties – problem owner, data scientists, other personas affected by the decision-making process – is required.

The next step is the collection and exploration of data. Data is gathered from the real or virtual world and activities must be conducted in order to get familiar with the data, identify trends, patterns and potential issues related to quality or completeness. Later, data must be prepared for the next modelling phase²⁰⁷. In the data preparation phase, data must be cleaned, organized and integrated in order to achieve completeness, accuracy and consistency. This means making data compatible, handling missing values²⁰⁸, detecting outliers which can interfere with the analysis, etc. Moreover, full compliance with the relevant legislation must be ensured. Feature engineering or feature learning (or representation learning)²⁰⁹ are alternative techniques that are crucial in the data preparation phase in order to identify or derive (new) features that are considered relevant in the datasets. Extreme caution is necessary in this phase, as bias may depend on the choices made therein.

In the modelling phase, the AI model is built or trained²¹⁰. As mentioned above, depending on the problem to be solved, the solution to be obtained and the available clean datasets, the computer scientists try out different techniques and algorithms to identify the ones that work best empirically. Once the learning algorithm is identified the AI model is trained and parameters are calibrated (so called tuning phase). Once the AI model is built or trained, evaluating its performance is fundamental. It requires to review the steps executed up to this moment and generally the model is evaluated based on new, unseen data. At this stage, issues can arise because of errors in the previous phases, such as in the specification of the problem, the collection and preparation of the data or in the modelling phase. In the event of problems, it is necessary to return to the relevant phase. Conversely, in the event of positive validation, the deployment phase begins, and the AI system is integrated into the operative process. At this stage, as in previous stages, it is crucial to check compatibility and ensure regulatory compliance. After the deployment phase, the AI system must be subject to constant monitoring in order to identify potential problems and make the appropriate adjustments. As recalled by Guidelines adopted by the Italian Agency for digital Italy (hereinafter “AgID”) the organisation must carry out maintenance activities, re-training and optimisation of the model, constantly assessing the achievement

²⁰⁷ I deal with the topic of data in Chapter 4.

²⁰⁸ Choices must be made. What is the best solution? Setting missing values to zero? Dropping records with missing values? Filling missing value?.

²⁰⁹ Y. BENGIO, A. COURVILLE, P. VINCENT, *Representation Learning: A Review and New Perspectives*, 2012-2014. Accessible at: <https://doi.org/10.48550/arXiv.1206.5538>.

²¹⁰ For a trustworthy modelling phase, albeit specifically for machine learning, Varshney suggests following three main phases: 1) pre-processing the training data, 2) training the model with appropriate ML algorithm, and 3) post-processing the model’s output predictions. VARSHNEY, *Trustworthy Machine Learning*, cited, 19-21

CHAPTER 1

of the objectives and operational effectiveness²¹¹. If malfunctions are detected or when it is deemed that the AI system is no longer needed, it must be withdrawn or deactivated.

The concept of *lifecycle* is extremely relevant from a legal perspective, especially given the newly adopted AI Act and the obligations set out therein, which must be ensured throughout the entire AI lifecycle. Interestingly, the AI Act does not define this concept but refers to it several times²¹². Indeed, the risk management system envisioned by Article 9 of the AI Act should be considered as a continuous iterative process planned and run throughout the entire lifecycle of the high-risk AI system. Similarly, right to privacy and to protection of personal data must be guaranteed throughout the entire lifecycle of the AI system and similarly accuracy, robustness and cybersecurity must be ensured throughout the entire AI lifecycle. These requirements are prescribed by the AI Act to mitigate risks inherent to AI system utilization, which may emerge across different phases of the AI lifecycle²¹³.

²¹¹ AgID, *Guidelines*, cited, 16.

²¹² Reference goes to the lifecycle of high-risk AI system (Recital 65, 69, 73, 74; Article 9 par. 2, 15 par.1, 40 par. 2) or to the lifecycle of the model when general-purpose AI models come into consideration (Recital 110, 114, 115). ISO/IEC 22989:2022 defines the lifecycle in relation to AI as an «evolution of a system, product, service, project or other human-made entity, from conception through retirement». ISO/IEC 22989:2022, paragraph 3.1.22.

²¹³ Cfr. Chapter 3, Part I, Paragraph 2.2.

Part II: Theoretical coordinates on law enforcement

1. The complex and fragmented landscape of law enforcement in the European Union

This Thesis examines the use of AI by law enforcement authorities. Therefore, it is essential to articulate a clear and contextually grounded definition of law enforcement and to delineate its boundaries, in order to ensure the correct application of the relevant provisions of the AI Act. While the term might appear straightforward, it encompasses a diverse array of institutions, roles, and practices that vary significantly across jurisdictions. Law enforcement typically refers to the public function carried out by public authorities who are empowered to uphold and enforce laws, maintain public order, prevent and investigate crime, and ensure community security. However, this definition becomes more complex when one considers the growing involvement of non-traditional actors, such as private entities or bodies, intelligence services, and transnational policing networks in activities that intersect with law enforcement functions. Moreover, the emergence of AI technologies has begun to blur conventional boundaries, redistributing responsibilities between human officers, automated systems, and third-party vendors. As I will discuss throughout this Chapter, understanding what constitutes law enforcement is not merely a terminological exercise but a foundational step in critically assessing the implications of AI integration within these bodies.

At the European level, the concept of law enforcement is set out in the Directive (EU) 2016/680 (hereinafter “Law Enforcement Directive”, “Directive”, or “LED”)²¹⁴ establishing the Directive’s

²¹⁴ Before the LED, there was a lack of a harmonised, coherent framework for data protection within the Area of Freedom, Security and Justice. The Council Framework Decision 2008/977/JHA was adopted on 27 November 2008. It was introduced to setting common data protection standards at a horizontal level in this area. Following the Framework Decision, the LED represented a major step forward in establishing a comprehensive data protection regime at the European level in the police and criminal justice sectors. The Framework Decision was indeed limited in scope and inadequate in terms of data protection standards and was inadequate in light of article 16 TFEU. Directive 2016/680 merely provides for minimum harmonisation (Article 1(3)), as Member States can provide higher data protection safeguards under national law. The research covering legislation on the processing of personal data in the law enforcement sector is very little compared to the one focused on the GDPR. Among the works on the topic see J.A. CANNATA, M.M. CARUANA, *Recommendation R (87) 15 – Twenty-five years down the line*, Report, February 2014. S. BRAUM, V. COVOLO, *From proven fragmentation to guaranteed data protection within the virtual criminal law enforcement area: A report on personal data protection within the framework of police and judicial cooperation in criminal matters* in K. LIGETI (Ed.), *Toward a Prosecutor for the European Union – Volume 1: A Comparative Analysis*, Hart Publishing, 2012. T. MARQUENIE, *The Police and Criminal Justice Authorities Directive: Data protection standards and impact on the legal framework*, in *Computer law & Security review*, 33(3), 2017, 324–340. M.M. CARUANA, *The reform of the EU data protection framework in the context of the police and criminal justice sector: harmonisation, scope, oversight and enforcement*, in *International Review of Law, Computers & Technology*, 33(3), 2017, 249–270. M. BREWCZYŃSKA, *A critical reflection on the material scope of the application of the Law Enforcement Directive and its boundaries with the General Data Protection Regulation* in E. KOSTA, R. LEENES, I. KAMARA (Eds.), *Research handbook on EU data protection law*, Edward Elgar Publishing Ltd, 2022, 91–114; J. SAJFERT, T. QUINTEL, *Data Protection Directive (EU) 2016/680 for Police and Criminal Justice Authorities*, 2017. Available at SSRN: <https://ssrn.com/abstract=3285873> or <http://dx.doi.org/10.2139/ssrn.3285873>. P. DE HERT, J.

scope of application. More precisely, law enforcement refers to activities undertaken by “*any public authority competent for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including safeguarding against and preventing threats to public security*”²¹⁵.

Therefore, the data controller responsible for the processing of personal data must be a “competent authority”²¹⁶ (personal scope) and the processing of personal data must be for “law enforcement purposes”²¹⁷ (material scope). This definition is notably expansive, extending not only to traditional law enforcement authorities, but also to any other body or entity that is entrusted by the legislation of a Member State with exercising public authority and powers for the same purposes²¹⁸. The Directive thus only applies when personal data is processed by public authorities or private entities that have been entrusted with law enforcement purposes by the legal system²¹⁹. Furthermore, the definition of law enforcement functions as a boundary marker that excludes the applicability of the General Data Protection Regulation (hereinafter “GDPR”)²²⁰, which explicitly does not apply in contexts involving law enforcement purposes²²¹. In certain situations, such as those involving migration and border

SAJFERT, *The fundamental right to personal data protection in criminal investigations and proceedings: framing big data policing through the purpose limitation and data minimisation principles of the Directive (EU) 2016/680*, Brussels Privacy Hub, 7(31), 2021. Available at: <https://ssrn.com/abstract=4016491> or <http://dx.doi.org/10.2139/ssrn.4016491>. M.R. LEISER, B. CUSTERS, *The Law Enforcement Directive: Conceptual Issues of EU Directive 2016/680*, in *European Data Protection Law Review*, 5(3), 2019, 367-378. S. AIDINLIS, D. BARNARD-WILLS, L. COCHRANE, K. GARSTKA, A. GURZAWSKA, J. HUGHES, *Between GDPR and Law Enforcement Directive in Security Research: The Use of Personal Data by Law Enforcement Authorities*, in *European Journal of Law and Technology*, 15(3), 2024. The most comprehensive work on the data protection legislation in law enforcement contexts is E. KOSTA, F. BOEHM (Eds.), *The EU Law Enforcement Directive (LED): A Commentary*, Oxford University Press, 2024.

²¹⁵ LED, Article 3(7).

²¹⁶ LED, Article 1(1).

²¹⁷ LED, Article 2(1).

²¹⁸ LED, Recital 11 clarifies “*Such competent authorities may include not only public authorities such as the judicial authorities, the police or other law-enforcement authorities but also any other body or entity entrusted by Member State law to exercise public authority and public powers for the purposes of this Directive*”. In contrast, in the former Framework Decision 2008/977/JHA, Article 2(h), the definition of ‘competent authority’ only covered traditional law enforcement authorities. “*In the Commission’s view, ‘competent authorities’ as defined by the LED are either organs of the State or private bodies, on which the law confers special powers beyond those which result from the normal rules applicable in relations between individuals and/or by the possibility of exercising the power of coercion*”. European Commission, *Communication from the Commission to the European Parliament and the Council. First report on application and functioning of the Data Protection Law Enforcement Directive (EU) 2016/680 (‘LED’)*, July 2022. Available at: https://commission.europa.eu/publications/first-report-application-and-functioning-data-protection-law-enforcement-directive-eu-2016680-led_en. For a discussion on the definition of “competent authorities” within the LED see L. TOSONI, L.A. BYGRAVE, *Article 3. Definitions*, in E. KOSTA, F. BOEHM (Eds.), *The Law Enforcement Directive: A Commentary*, cited, 99-104.

²¹⁹ Therefore, if personal data are processed by law enforcement authorities for non-law enforcement purposes, such as for human resource activities, the LED does not apply.

²²⁰ The definition of “competent authorities” in the LED must be applied, by analogy, to Article 2(2)(d) of Regulation 2016/679. See also CJEU, Judgment of 22 June 2021, *Latvijas Republikas Saeima (Penalty points)*, C-439/19, EU:C:2021:504, paragraph 69.

²²¹ Article 2(2)(d) GDPR provides that the Regulation does not apply to the processing of personal data “*by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security*”.

control, however, it is not always clear whether the LED or the GDPR applies. The applicable framework depends on the specific purpose of the personal data processing, which in turn determines the obligations that the relevant authorities must fulfil²²².

Since the AI Act reproduces the definition almost verbatim²²³, this alignment suggests a deliberate effort of the European legislator to create regulatory coherence between data protection and AI governance in the law enforcement domain. Therefore, it becomes particularly important to analyse not only the text of the LED itself but also the associated recitals, relevant case law, and academic literature that have interpreted and applied this definition over time. Such analysis can provide valuable insights into how the term is likely to be understood and operationalised within the AI Act, especially in relation to the provisions that provide some specificities when AI systems are deployed by or on behalf of law enforcement bodies for law enforcement purposes²²⁴.

The subsequent analysis underscores the enduring lack of clarity surrounding certain core concepts of the LED. This ambiguity poses not only practical difficulties in delineating the respective material scopes of the LED and the GDPR but also raises concerns in light of the AI Act. Indeed, the latter expressly incorporates, in Articles 3(45) and 3(46), the definition set forth in Article 3(7) of the LED²²⁵. From a doctrinal perspective, such legislative alignment reflects a deliberate attempt to ensure consistency and continuity across the Union's regulatory architecture. As a result, interpretative approaches developed in relation to the LED necessarily inform the exegesis of the AI Act. Nevertheless, this very alignment risks perpetuating the conceptual indeterminacy already identified, thereby undermining the principles of legal certainty and foreseeability that are foundational to EU law. In particular, the precise delineation of “*law enforcement authorities*” and “*law enforcement purposes*” constitutes the *conditio sine qua non* for the correct application of the AI Act's special regime on law enforcement. Indeed, as I will analyse in Chapter 3, the AI Act introduces a differentiated regulatory regime for AI systems used in the context of law enforcement, including prohibitions on certain practices and specific obligations concerning high-risk AI systems. Hence, the preliminary hermeneutic task in applying the Regulation is to establish whether, in a given case, an authority acting as provider or deployer of AI systems may properly be characterised as a law enforcement authority operating for law enforcement purposes.

²²² SAJFERT, QUINTEL, *Data Protection Directive (EU) 2016/680 for Police and Criminal Justice Authorities*, cited, 3. In some cases, personal data of migrants may be processed to prosecute illegal border crossings, in which case the LED applies. In other cases, their personal data may be processed to apply for asylum, in which case the GDPR applies.

²²³ AI Act, Article 3(45) “‘*law enforcement*’ means activities carried out by law enforcement authorities or on their behalf for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including safeguarding against and preventing threats to public security”.

²²⁴ European Commission, *Guidelines on Prohibited Artificial Intelligence Practices*, paragraph 320.

²²⁵ *Ibidem*.

1.1. Defining criminal offence, criminal penalty and public security

The definition of law enforcement is structured around three main concepts, namely those of “criminal offences”, “criminal penalties”, and “public security”. These concepts do not merely describe different legal categories but also reflect “*different points in time and different activities*”²²⁶ of the law enforcement function. In particular, investigating criminal offences and executing criminal penalties represent different stages in the State’s reactive response *after* a crime has been committed. More precisely, both the “investigation” and “detection” of crime refer to the pre-trial phase, which involves identifying suspects and gathering information that may become evidence to support or refute allegations. This stage prepares the ground either for pressing charges or closing the case. “Prosecution” typically begins when the prosecutor formally files charges in court, marking the start of the trial. The “execution of criminal penalties” follows the trial phase and involves carrying out the imposed sentence. These phases are traditionally governed by national criminal procedural law, which can differ greatly between Member States²²⁷. By contrast, “prevention” involves the proactive measures undertaken by competent authorities to avert crime or safeguard public order and security *before* any offence has been committed or formally brought under investigation²²⁸. One of the current strategies of crime prevention, for example, concerns the deployment of predictive policing, that I will discuss in Chapter 2. While preventive activities may occasionally overlap with criminal law, they are more commonly associated with administrative law, particularly when measures are implemented

²²⁶ Opinion of Mr Campos Sánchez-Bordona, CJEU, Judgement of 8 December 2021, *Inspektor V Inspektorata Kam Visshia Sadeben Savet*, Case C-180/21, ECLI:EU:C:2022:967, paragraph 52. In particular, the Advocate Generale identifies three purposes carried out by law enforcement authorities: i) the purpose of “prevention”, ii) the purpose of “investigation” of criminal acts and lastly iii) the purpose of “execution of criminal penalties”.

²²⁷ P. VOGIAZOGLOU, T. MARQUENIE, *Assessment of the implementation of the Law Enforcement Directive*, Study requested by the European Parliament’s Committee on Civil Liberties, Justice and Home Affairs, November 2022, 18. Available at: [https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU\(2022\)740209](https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2022)740209). The LED excludes court proceedings and bodies acting in their judicial capacity from certain data protection requirements. On this aspect broad discretion is left to the Member States in regulating data processing for courts. LED, Recital 20 states that “*This Directive does not preclude Member States from specifying processing operations and processing procedures in national rules on criminal procedures in relation to the processing of personal data by courts and other judicial authorities, in particular as regards personal data contained in a judicial decision or in records in relation to criminal proceedings*”.

²²⁸ LED, Recital 12 “*The activities carried out by the police or other law-enforcement authorities are focused mainly on the prevention, investigation, detection or prosecution of criminal offences, including police activities without prior knowledge if an incident is a criminal offence or not. Such activities can also include the exercise of authority by taking coercive measures such as police activities at demonstrations, major sporting events and riots. They also include maintaining law and order as a task conferred on the police or other law-enforcement authorities where necessary to safeguard against and prevent threats to public security and to fundamental interests of the society protected by law which may lead to a criminal offence*”. “*Crime prevention refers to efforts to prevent crime or criminal offending in the first instance – before the act has been committed*”. D.P. FARRINGTON, B.C. WELSH, *Crime Prevention and Public Policy*, in D.P. FARRINGTON, B.C. WEL (Eds.) *The Oxford Handbook of Crime Prevention*, Oxford University Press, 2012, 3-19.

without reference to a specific suspect or offence. For example, surveillance in high-risk areas or gathering of intelligence on potential public security threats falls into this category.

The notion of law enforcement, therefore, does not include only classic investigative activities – which represent the core focus of law enforcement – but captures also pre-crime intelligence and post-sentencing supervision. Each of these three purposes engages different types of authorities, activities, and legal regimes. Notably, whereas investigation and execution are firmly anchored in criminal procedural law, the preventive function often extends into the domain of administrative law.

Currently, there is no consensus on the definition of “criminal offence” across EU Member States, which inevitably creates inconsistencies across jurisdictions. To avoid legal uncertainty, therefore, the European legislator, for the purposes of the LED, opted for the development of an autonomous concept which should not be defined solely by the legal framework of individual Member States, but rather interpreted through the lens of EU law itself²²⁹. In order to define an offence as criminal, reference should be made to the substantive criteria developed through the case law of the Court of Justice (hereinafter “CJEU”), which have been particularly influenced by the *Engel* criteria established by the European Court of Human Rights (hereinafter “ECtHR”)²³⁰. According to the *Engel* criteria, which represent a fundamental reference to distinguish between criminal and administrative offences, the classification of an act as a criminal offence depends on its legal label under national law (first criterion), on its essential nature (second criterion) and the severity of the penalty that could be imposed on to the person concerned (third criterion). Despite the European Union’s efforts – and partly because the first criterion remains closely tied to national classifications – Member States continue to interpret what constitutes a criminal offence by relying on the definitions provided in their own national legal system²³¹. This is mainly due to the fact that concepts used in criminal law were born and evolved in line and within national political, social and constitutional systems.

²²⁹ LED, Recital 13. About the doctrine of autonomous concepts in the European context see V. MITSILEGAS, *Autonomous Concepts, Diversity Management and Mutual Trust in Europe’s Area of Criminal Justice*, in *Common Market Law Review*, 57(1), 2020, 45-78.

²³⁰ TOSONI, BYGRAVE, *Article 3. Definitions*, cited, 79-132. Commission Expert Group, Minutes of the meeting of the Commission expert group on the Regulation (EU) 2016/679 and Directive (EU) 2016/680, 7 November 2016, paragraph 1. The *Engel* criteria were established in ECtHR, Plenary, Judgement of 8 June 1976, *Case of Engel and others v. the Netherlands*, Application no. 5100/71; 5101/71; 5102/71; 5354/72; 5370/72. See also ECtHR, Sec. II, Judgement of 4 March 2014, *Case of Grande Stevens v. Italy*, Application No. 18640/10. For a critical analysis of the *Engel* criteria see B. BAHCECI, *Redefining the Concept of Penalty in the Case-law of the European Court of Human Rights*, in *European Public Law*, 26(4), 2020, 867 – 888.

²³¹ The Commission Expert Group on the GDPR and LED initially responded affirmatively when the question was raised about using those indicators to determine the scope of the LED’s application. However, as discussions progressed, the approach shifted, and a few months later, the European Commission’s Legal Service argued that, in transposing the LED, Member States may base their interpretation of a criminal offence on definitions provided in their own national legal systems. For the different considerations see Commission Expert Group 2016, cited, paragraph 1 and Commission Expert Group, Minutes of the meeting of the Commission expert group on the Regulation (EU) 2016/679 and Directive (EU) 2016/680, 4 May 2017, paragraph 1.

The lack of concrete steps towards harmonisation of such concepts means that there are still differences in national definitions of criminal offences, which could result in uncertainties surrounding the notion of law enforcement itself. In relation to data protection legislation, some Member States still consider the boundary between the GDPR and the LED to be a challenging issue, due to the classification of minor offences or misdemeanours as part of national criminal law, even though the responsibility for detecting and investigating such offences may lie with bodies that are not traditional law enforcement authorities²³². Moreover, in some cases, administrative offences explicitly fall within the scope of the national legislation that transposes the LED. In the case of Latvia, for instance, administrative breaches within multiple areas (employment law, traffic, customs, environmental law, etc.), whose responsibility is of administrative authorities tasked with detecting them and imposing penalties, are considered to fall within the scope of application of the LED. Similarly, Germany, both at federal and Länder level, has brought administrative offences within the scope of application of the LED²³³.

Similar considerations can be made for “criminal penalties”, which follow the identification and prosecution of a criminal offence. As with criminal offences, they must be assessed based on their substance, such as the degree of severity, deterrence or stigmatisation, rather than on how they are categorised within the national legal system²³⁴. Similar to the case of criminal offences, the lack of concrete harmonisation could lead to blur the boundaries of the definition of law enforcement.

Comparable, and indeed even greater, uncertainties stem from the provision’s explicit inclusion of “*the safeguarding and prevention of threats to public security*” within the notion of law enforcement²³⁵. This addition, wanted by the Council during the proposal phase, was widely criticised, due to the uncertainty traditionally surrounding the concept of “public security”, which remains vague and open to interpretation²³⁶. Some scholars, for example, argue that the use of the word “*including*” before this addition implies that such tasks must remain closely linked to the “*prevention [or] detection [...] of*

²³² Commission Expert Group 2016, cited, paragraph 1.

²³³ The BDSG clearly states that Article 3§45(1) BDSG related to “competent authorities” “*shall apply to the processing of personal data by public bodies competent for the prevention, investigation, detection or prosecution of criminal or administrative offences or the execution of criminal or administrative penalties, as far as they process data for the purpose of carrying out these tasks. The public bodies shall be regarded in that case as controllers. The prevention of criminal offences as referred to in the first sentence shall include protection against and prevention of threats to public security. The first and second sentences shall also apply to those public bodies responsible for executing penalties, measures as referred to in Section 11 (1) no. 8 of the Criminal Code, educational or disciplinary measures as referred to in the Juvenile Court Act or fines. As far as this Part contains provisions for processors, it shall also apply to them*

²³⁴ The Engel criteria apply also to the concept of “criminal penalties”. CJEU, Grand Chamber, Judgement of 26 February 2013, *Åklagaren v. Hans Åkerberg Fransson*, Case C-617/10, paragraph 35.

²³⁵ This addition “*can receive different interpretations and does not provide for a clear delimitation of the tasks of the police within the scope of the Directive*”. EDPS, *Opinion 6/2015. A further step towards comprehensive EU data protection*, 28 October 2015, 5-6.

²³⁶ Criticism due to the lack of clarity of the term “public security” was stressed by the EDPS, cited, 6; Article 29 Working Party, *Opinion 03/2015 on the draft directive on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data*, WP233, 1 December 2015, 5.

criminal offences” as outlined at the beginning of Article 3(7)(a). From this perspective, the reference to public security should still be interpreted as tied to actual or potential criminal offences in order to fall within the scope of the definition²³⁷. In contrast, during the legislative process of the LED, some Member States advocated for a broader interpretation, contending that the LED should extend to the general field of threat prevention by the police, even when the legal interests at stake are not necessarily protected by criminal law²³⁸. In this author’s assessment, the latter interpretation seems to align more closely with the Council of Europe’s “*Recommendation on the Use of Personal Data in the Police Sector*”²³⁹, which clarifies that the term “*police purposes*” encompasses not only the prevention and suppression of criminal offences but also the maintenance of public order²⁴⁰.

The inclusion of the protection of public security within the notion of law enforcement has important implications. While the prevention of threats to public security may involve the use of police powers – such as surveillance, identity checks, or the dispersal of crowds – these actions are often not linked to a specific criminal offence or a criminal investigation. Instead, they are rooted in administrative objectives aimed at maintaining public order, managing risks, or generally ensuring the safety of individuals and communities²⁴¹. Accordingly, should the latter interpretation be adopted, whereby the notion of law enforcement is construed as encompassing considerations of public security and public order, the GDPR shall not apply; rather, the LED will govern, together with the special regime it establishes. This conclusion finds support in Recital 11 of the LED, which enumerates examples of cases in which the Directive applies, such as police activities at demonstrations, major sporting events

²³⁷ TOSONI, BYGRAVE, *Article 3. Definitions*, cited, 101.

²³⁸ Germany agrees with Romania’s analysis in that the area of threat prevention by the police (“administrative police”) should not be included in the scope of the GDPR. Instead, Germany finds it necessary for subject-related and legal reasons for the LED to cover general threat prevention by the police, *i.e.* regardless of whether legal interests protected by criminal law are at issue. “*The provisions of this directive may also be intended to govern processing carried out in the context of activities that do not specifically fall within the criminal sphere but which relate to police activities carried out prior to the commission of a criminal offence*”. CNIL, “*Law enforcement Directive*”: *What Are We Talking About?*, 2 June 2021. Available at: <https://www.cnil.fr/en/law-enforcement-directive-what-are-we-talking-about>.

²³⁹ Council of Europe, *Recommendation No. R(87)15 of the Committee of Ministers to Member States regulating the use of personal data in the police sector*, 17 September 1987.

²⁴⁰ LED, Recital 12: “*The activities carried out by the police or other law-enforcement authorities are focused mainly on the prevention, investigation, detection or prosecution of criminal offences, including police activities without prior knowledge if an incident is a criminal offence or not. Such activities can also include the exercise of authority by taking coercive measures such as police activities at demonstrations, major sporting events and riots. They also include maintaining law and order as a task conferred on the police or other law-enforcement authorities where necessary to safeguard against and prevent threats to public security and to fundamental interests of the society protected by law which may lead to a criminal offence. Member States may entrust competent authorities with other tasks which are not necessarily carried out for the purposes of the prevention, investigation, detection or prosecution of criminal offences, including the safeguarding against and the prevention of threats to public security, so that the processing of personal data for those other purposes, in so far as it is within the scope of Union law, falls within the scope of Regulation (EU) 2016/679*”.

²⁴¹ Council of the European Union, Report, 30 October 2013, 2012/0010 (COD), 9: “*Germany agrees with Romania’s analysis in that the area of threat prevention by the police (“administrative police”) should not be included in the scope of the General Data Protection Regulation. Instead, Germany finds it necessary for subject-related and legal reasons for the Directive to cover general threat prevention by the police, i.e. regardless of whether legal interests protected by criminal law are at issue*”.

and riots, or more in general, police activities maintaining law and order, which are all activities that do not fall within the criminal sphere.

This functional overlap blurs the boundary between the two legal domains, the administrative one and the criminal one, raising questions about the appropriate legal bases, safeguards, and oversight mechanisms applicable to such data processing activities. Nonetheless, the Council's decision to include the reference to the protection and prevention of threats to public security is justified by both a strategic intent and a practical need: to ensure that the full spectrum of police duties would fall under the scope of the LED rather than the GDPR²⁴², ensuring that all core functions of police authorities are regulated under the same legal instrument, avoiding legal uncertainties and overlapping.

This reflects the current operational realities of policing, where the focus extends beyond simply responding to criminal activities after they occur. Modern law enforcement now emphasises proactive measures aimed at preventing crimes before they happen. This shift involves a range of strategies, including intelligence gathering, surveillance, and the use of predictive analytics, whose aim is to disrupt criminal behaviour before it escalates, ultimately reducing crime rates and enhancing public security. Consequently, such addition reflects the evolving nature of policing, where administrative and criminal functions are increasingly intertwined, but it also requires careful legal and operational distinctions to prevent ambiguity, safeguard individual rights, and ensure proper accountability.

Moving to the AI Act, the fact that the definition of law enforcement enshrined in the LED extends beyond the investigation of criminal offences to also encompass their prevention and detection, as well as activities aimed at safeguarding public security has important consequences. Adopting such an interpretation means that systems used for these purposes fall within the specific provisions governing AI systems deployed for law enforcement activities. If the safeguarding of public security is subsumed under law enforcement, then AI-driven practices such as predictive policing, biometric identification in public spaces for the purposes of police operations during demonstrations, major sporting events, or riots would automatically trigger the application of the specific provisions foreseen for law enforcement activities. By contrast, if public security were to be construed as a distinct policy domain falling outside the scope of law enforcement, the deployment of AI systems in this field would be governed solely by the general provisions of the AI Act, thereby limiting the applicability of the sector's specific provisions. Such an interpretation could result in reduced protection, whereby the use

²⁴² Council of the European Union, Report, cited, 101: “[Romania] would like to draw attention that the unclear delimitation between the Directive and Regulation will have major negative impact on the law enforcement authorities. Practical activities of the Police include the maintaining and ensuring the public order, during which the processing of personal data is a very important part. Leaving these activities under the regulation would practically block the possibility to legally process personal data of persons when fulfilling duties by the police officers. It is compulsory that the maintaining and ensuring the public order be regulated by the directive”.

of highly intrusive technologies escapes the more rigorous safeguards despite their potentially far-reaching implications for fundamental rights.

In light of the European Union's explicit objective with the AI Act to enhance the protection of individuals and communities against the risks posed by AI systems, the former interpretation appears more consistent with the Regulation's stated objective. By contrast, a restrictive understanding of law enforcement purposes confined only to activities directly linked to a specific criminal offence or suspect would unduly curtail the scope of protection envisaged by the European legislator. It should instead be recognised that AI systems may be deployed by law enforcement authorities in preventive or security-oriented contexts, even in the absence of an identified offence or suspect. The fact that such practices may eventually contribute to the detection or prosecution of crimes does not diminish their capacity to interfere with fundamental rights already at earlier, pre-investigative stages. As such, a purposive and rights-oriented interpretation of "law enforcement purposes" is necessary to avoid undermining the effectiveness of the safeguards established by the AI Act. The *Guidelines on Prohibited Artificial Intelligence Practices* (hereinafter "Guidelines") confirm this interpretation, clarifying that law enforcement purposes include "*the investigation, detection, and prosecution of criminal offences*" as well as "*activities in relation to the prevention of criminal offences, including safeguarding against and the prevention of threats to public security, before any crime has actually been committed*"²⁴³. They also recall verbatim Article 3(7) of the LED, which states that, for instance, "*the police may take 'coercive measures at demonstrations, major sporting events or riots' in the context of crime prevention*".

1.2. National security and public security

The concept of law enforcement is further complicated by the distinction between activities carried out in the interest of *public security* and those in the interest of *national security*²⁴⁴. Traditionally, the distinction between the two types of security lies with the respective authorities responsible for them: intelligence services for national security and law enforcement authorities for public security. However, this distinction is no longer determinant due to the increased cooperation between the two organisations and their overlapping activities.

²⁴³ European Commission, *Guidelines on Prohibited Artificial Intelligence Practices*, cited, paragraph 321.

²⁴⁴ P. VOGIATZOGLU, S. FANTIN, *National and Public Security Within and Beyond the Police Directive*, in A. VEDDER, J. SCHROERS, C. DUCUING, P. VALCKE (Eds). *Security and Law. Legal and Ethical Aspects of Public Security, Cyber Security and Critical Infrastructure Security*, Intersentia, 2019, 123-156. Available at SSRN: <https://ssrn.com/abstract=3466821>.

The concept of “*national security*” is closely tied to the notions of State sovereignty and national identity and encompasses measures and activities undertaken by a State to safeguard its fundamental political, territorial, and institutional integrity²⁴⁵. Whenever the sovereignty or identity of a State is perceived to be at risk, whether from external aggression, foreign interference or other forms of coercion, such threats are usually considered to be a matter of national security, which therefore serves as a broad framework for protecting the attributes that are essential to the State and its position in the global order²⁴⁶.

However, defining the concept of national security from a legal point of view is extremely complex, as it is inherently ambiguous, due to different understandings within international, European and national legal frameworks. At the national level, every country has its specific conceptualization of national security²⁴⁷, with certain Member States having adopted an explicit legal definition within their domestic legal frameworks. Conversely, at the international level, no proper definition of national security can be found. Nevertheless, many international treaties explicitly recognise national security as a legitimate basis for derogating from certain obligations²⁴⁸. While national security exemption clauses recognise that States may need to take extraordinary measures to protect national security, treaties generally require that such measures are taken in respect of the principle of necessity and proportionality²⁴⁹. Therefore, courts and treaty bodies increasingly scrutinise claims to ensure that

²⁴⁵ An interesting report which specifically deals, among other aspects, with the data processing for national security purposes is J. RIJPMAN ET AL. (Eds), *The New EU Data Protection Regime: Setting Global Standards for the Right to Personal Data Protection*, the XXIX FIDE Congress in The Hague 2020 Congress Publications, vol. 2, Eleven International Publishing, 2020.

²⁴⁶ For a detailed analysis of the concept of national security see A. STROZZI (Ed.), *Sicurezza nazionale. Poteri, conflitti, informazioni*, Luiss University Press, 2024.

²⁴⁷ For an overview on the different understandings of national security see Council of Bars & Law Societies of Europe, *Recommendations on the protection of fundamental rights in the context of 'national security'*, 2019. Available at: https://www.ccbe.eu/fileadmin/speciality_distribution/public/documents/SURVEILLANCE/SVL_Guides_recommandations/EN_SVL_20190329_CCBE-Recommendations-on-the-protection-of-fundamental-rights-in-the-context-of-national-security.pdf.

²⁴⁸ Among the many works on the topic see S. ROSE-ACKERMAN, B. BILLA *Treaties and National Security*, in *International Law And Politics*, 40, 2008, 437–496; D. EISENHUT, *Sovereignty, National Security and International Treaty Law. The Standard of Review of International Courts and Tribunals with regard to 'Security Exceptions'*, in *Archiv Des Völkerrechts*, 48(4), 2010, 431–466. Available at: <http://www.jstor.org/stable/25782611>. To give some examples, the General Agreement on Tariffs and Trade (GATT) includes (national) security exceptions (Article XXI) and the International Covenant on Civil and Political Rights (ICCPR) mentions national security as an exception or restriction to the exercise of several rights (Article 12, 13, 14, 19, 21 and 22). Similarly, the ECHR considers national security as an exception to the full enjoyment of the rights enshrined therein. National security is often invoked in the context of restrictions on human rights, but international treaties require balancing security concerns with fundamental rights. The exception of national security can be invoked solely where explicitly mentioned. The national security exception cannot be invoked, for example, in relation to the prohibition against torture or slavery or the prohibition against unlawful punishment. In other cases, the exception must be balanced with compelling interests. In relation to data protection, Convention 108+ considers national security as a possible exception within specific limits (Article 11).

²⁴⁹ International law places limits on the use of national security justifications: the respect of proportionality and necessity principles require that measures taken for national security must be strictly necessary and proportional to the threat; non-discrimination requires that national security measures should not be used to disguise discriminatory or protectionist

national security exception clauses are not being used arbitrarily or excessively as a pretext to avoid treaty obligations or undermine international norms²⁵⁰.

As in the international context, there is no definition of national security in any legislative act in the European context either. This reflects the European legislator's deliberate choice to leave further contextualisation to the CJEU, since the concept of security (national and public) is generally dynamic and flexible, often evolving or adapting to different circumstances and legal scenarios. However, in many cases, legislative acts have considered national security to be exempt from their provisions²⁵¹. The legal basis on which such exceptions are based is Article 4(2) of the Treaty on the European Union (hereinafter "TEU") which clearly states that "[n]ational security remains the sole responsibility of each Member State"²⁵². This provision acknowledges the constitutional balance between EU integration and national sovereignty, recognising that certain core State functions, particularly those relating to national identity and sovereignty, are beyond the remit of EU law. Since national identity and sovereignty are closely linked to national security, Member States remain the sole holders of prerogatives in such matters.

Academic literature has emphasised that, despite the presence of a clause enshrined in Article 4(2) TEU, Member States cannot unilaterally decide to override obligations arising from EU law²⁵³. Similarly, the CJEU has always interpreted national security exceptions narrowly²⁵⁴. In ruling *La*

policies; provision of oversight mechanisms: Treaties often include monitoring bodies that can assess whether national security claims are justified.

²⁵⁰ When a State invokes a national security exception the CJEU intervene with its judicial scrutiny. Among others, see ECtHR, Judgement of 6 September 1978, *Case of Klass and others v. Germany*, Application no. 5029/71; ECtHR, Judgement of 21 October 2013, *Case of Janowiec and others v. Russia*, Applications nos. 55508/07 and 29520/09.

²⁵¹ See, among others, the e-Privacy Directive. Article 15 stipulates that "*Member States may adopt legislative measures to restrict the scope of [enumerated] rights and obligations [...] when such restriction constitutes a necessary, appropriate and proportionate measure within a democratic society to safeguard [...] national security (i.e. State security)*". However, Article 15(1) of the Directive "*presupposes that the national legislative measures referred to therein fall within the scope of that directive, since it expressly authorises the Member States to adopt them only if the conditions laid down in the directive are met*". CJEU, Judgement of 6 October 2020, *Privacy International*, Case C-623/17, ECLI: EU: C: 2020: 790, paragraph 38.

²⁵² In secondary legislation, national security is also referred to as "*state security*". See Directive 2006/24/EC, Recital 4 and 9; Directive 2002/58/EC, Recital 11 and Article 15(1); Council Framework Decision 2008/977/JHA, Recital 5. By contrast, the European Union assumes an intergovernmental leadership role in developing a pan-European framework for security and defence. In this domain, the Union is entrusted with the task of safeguarding a high level of security, while operating within a mandate that reflects the principles of shared and attributed competences. Its actions extend to crime prevention, the facilitation of police cooperation, the promotion of mutual recognition of judicial decisions, and the approximation of criminal law provisions. At the same time, Article 73 TFEU makes clear that Member States retain full discretion to organise, under their own responsibility, any forms of cooperation or coordination between the competent departments of their respective administrations responsible for safeguarding national security.

²⁵³ In these terms M. CLAES, *The Primacy of EU Law in European and National Law*, in D. CHALMERS, A. ARNULL (Ed.), *The Oxford Handbook of European Union Law*, Oxford University Press, 2015, 178–211. B. GUASTAFERRO, *Beyond the Exceptionalism of Constitutional Conflicts: The Ordinary Functions of the Identity Clause*, in *Yearbook of European Law*, 31(1), 2012, 263–318.

²⁵⁴ CJEU, Judgment of 2 April 2020, *Commission v Poland, Hungary and the Czech Republic*, Joined Cases C-715/17, C-718/17 and C-719/17, ECLI:EU:C:2020:257, paragraph 143: "*In this connection, according to settled case-law of the Court of Justice, although it is for the Member States to adopt appropriate measures to ensure law and order on their*

Quadrature du Net and Others, the Court acknowledged that national security involves the protection of essential State functions and societal interests, including the prevention of acts that could destabilize constitutional, political, or economic structures²⁵⁵, emphasising that any derogation from EU law must be strictly limited to what is necessary for genuine national security concerns. In ruling *Privacy International*, the CJEU held that national security (or state security) exceptions apply when the State is involved in activities relating solely to State security matters²⁵⁶. Thus, national security can be invoked only in the case of activities of the State or State authorities which are “purely governmental”, that is, without the involvement of any private actor²⁵⁷. Accordingly, the mere fact that a decision concerns State security cannot automatically result in European Union law being inapplicable²⁵⁸.

The rulings cited above have played an important role in shaping the understanding of the notion of law enforcement across Member States. While law enforcement activities encompass matters of public security, they do not extend to those related to national security. Accordingly, the AI Act expressly clarifies that its provisions should not interfere with the national security prerogatives of the Member States, which therefore remain outside the scope of the Regulation²⁵⁹. The use of the verb ‘should’, however, allows Member States to determine for themselves whether such purposes are to remain

*territory and their internal and external security, it does not follow that such measures fall entirely outside the scope of European Union law. [...] It cannot be inferred that the Treaty contains an inherent general exception excluding all measures taken for reasons of law and order or public security from the scope of European Union law. The recognition of the existence of such an exception, regardless of the specific requirements laid down by the Treaty, might impair the binding nature of European Union law and its uniform application”. Similarly, CJEU, Judgement of 6 October 2020, *Privacy International v Secretary of State for Foreign and Commonwealth Affairs and Others*, Case C-623/17, ECLI: EU: C: 2020: 790, paragraph 44: “although it is for the Member States to define their essential security interests and to adopt appropriate measures to ensure their internal and external security, the mere fact that a national measure has been taken for the purpose of protecting national security cannot render EU law inapplicable and exempt the Member States from their obligation to comply with that law”.*

²⁵⁵ CJEU, Grand Chamber, Judgement of 6 October 2020, *La Quadrature du Net and Others v Premier ministre and Others*, Joined Cases C-511/18, C-512/18, C-520/18, ECLI:EU:C:2020:791, paragraph 135: “Article 4(2) TEU provides that national security remains the sole responsibility of each Member State. That responsibility corresponds to the primary interest in protecting the essential functions of the State and the fundamental interests of society and encompasses the prevention and punishment of activities capable of seriously destabilising the fundamental constitutional, political, economic or social structures of a country and, in particular, of directly threatening society, the population or the State itself, such as terrorist activities”.

²⁵⁶ CJEU, *Privacy International*, cited, paragraph 35 : « Article 1(3) of that directive excludes from its scope ‘activities of the State’ in specified fields, including activities in areas of criminal law and in the areas of public security, defence and State security, including the economic well-being of the State when the activities relate to State security matters”.

²⁵⁷ *Ibidem*. “The activities thus mentioned by way of example are, in any event, activities of the State or of State authorities and are unrelated to fields in which individuals are active (judgment of 2 October 2018, *Ministerio Fiscal*, C-207/16, EU:C:2018:788, paragraph 32 and the case-law cited)”. In the same terms, VOGIATZOGLU, MARQUENIE, *Assessment of the implementation of the Law Enforcement Directive*, cited, 23.

²⁵⁸ CJEU, Judgement of 4 June 2013, *ZZ v Secretary of State for the Home Department*, Case C 300/11, ECLI:EU:C:2012:563, paragraph 73 “First of all, it is clear that reliance by a Member State on interests of State security does not rule out the application of EU law and, in particular, the fundamental rights protected by the Charter. It is also not sufficient in itself to justify the decision not to inform the Union citizen, precisely and in full, of the grounds for an expulsion or exclusion decision taken against him by a Member State”.

²⁵⁹ AI Act, Recital 24 “As regards national security purposes, the exclusion is justified both by the fact that national security remains the sole responsibility of Member States in accordance with Article 4(2) TEU and by the specific nature and operational needs of national security activities and specific national rules applicable to those activities”.

excluded. For instance, in implementing the LED at the national level, certain Member States have chosen to explicitly exclude national security activities from its scope²⁶⁰, while others, conversely, have opted to extend the application of the LED to cover national security agencies, thereby voluntarily subjecting them to the Directive's data protection standards²⁶¹.

By contrast, public security pertains to the maintenance of internal order and security within the State's territory²⁶². Within the European Union context both the legislator and the CJEU have generally followed a broad interpretation of public security. Secondary legislation, for example, has at times defined public security as presupposing "*a genuine and sufficiently serious threat affecting one of the fundamental interests of society*"²⁶³. This includes, for instance, threats to the functioning of institutions and essential public services, the survival of the population, serious disruptions to foreign relations or the peaceful coexistence of nations, and risks to military interests. The CJEU has similarly adopted a broad understanding of public security, recognising that it may encompass both the internal and external security of a Member State. In its case law, the Court has affirmed that public security concerns may arise not only from traditional law enforcement issues but also from more systemic threats to state stability and societal functioning²⁶⁴. In *Tsakouridis* ruling, for example, the CJEU held that public security could be affected by "*a threat to the functioning of the institutions and essential public services and the survival of the population, as well as the risk of a serious disturbance to foreign relations or to peaceful coexistence of nations, or a risk to military interests*"²⁶⁵.

These broad understandings of public security appear to blur the line between matters of public security and those of national security, encompassing within the concept of public security elements that are

²⁶⁰ As far as France is concerned the French Data Protection Authority (CNIL) stressed that "*Processing carried out to ensure state security or national defense does not fall within the scope of the European Union and remains governed by the provisions of the French Data Protection Act alone*". CNIL, "*Law enforcement Directive*": *What Are We Talking About?*, cited.

²⁶¹ Belgium, Czech Republic, Finland, France, Hungary, Luxembourg, Netherlands and Slovenia apply data protection rules to processing operations by authorities such as intelligence services for the safeguarding of national security. See VOGIATZOGLU, MARQUENIE, *Assessment of the implementation of the Law Enforcement Directive*, cited, 24.

²⁶² Albeit not explicitly stated, "internal" and "public" security are seemingly used interchangeably.

²⁶³ Regulation (EU) 2018/1807 of The European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, Recital 19. See also CJEU, Preliminary ruling of 23 November 2010, *Land Baden-Württemberg v Panagiotis Tsakouridis*, Case C-145/09, ECLI:EU:C:2010:708.

²⁶⁴ CJEU, Judgement of 24 June 2015, *H. T. v Land Baden-Württemberg*, Case C-373/13, ECLI:EU:C:2015:413, paragraph 78.

²⁶⁵ CJEU, *Panagiotis Tsakouridis*, cited, paragraph 44. Similarly, CJEU, Judgement of 22 May 2012, *P.I. v Oberbürgermeisterin der Stadt Remscheid*, Case C-348/09, ECLI:EU:C:2012:300, paragraph 28. These definitions established by the CJEU have subsequently been incorporated into the Recitals of the Regulation on the free flow of non-personal data: "*the concept of public security, within the meaning of Article 52 TFEU and as interpreted by the Court of Justice, covers both the internal and external security of a Member State, as well as issues of public security, in order, in particular, to facilitate the investigation, detection and prosecution of criminal offences*"; it "*presupposes the existence of a genuine and sufficiently serious threat affecting one of the fundamental interests of society, such as a threat to the functioning of institutions and essential public services and the survival of the population, as well as the risk of a serious disturbance to foreign relations or the peaceful coexistence of nations, or a risk to military interests*". Regulation (EU) 2018/1807, Recital 19.

traditionally considered to be part of national security. Moreover, as with national security, public security is sometimes considered a possible exception to the entirety or specific provisions of EU law²⁶⁶. However, distinguishing between the two concepts is necessary, since the latter is explicitly excluded from the scope of EU law, as national security remains beyond the Union's competences²⁶⁷. In *La Quadrature du Net* ruling, the CJEU tried to articulate an EU-level understanding of national security, underling that it relates to “*the primary interest in protecting the essential functions of the State and the fundamental interests of society*”²⁶⁸. This encompasses the prevention and punishment of activities capable of seriously destabilizing the constitutional, political, economic, or social structures of a country, particularly those that pose a direct threat to society, the population, or the State itself, such as acts of terrorism. According to the Court, the distinction between national and public security lies in the nature and severity of the threat²⁶⁹. A threat to national security is characterised by its exceptional gravity and specific circumstances, setting it apart from the more general, ongoing risks, such as serious criminal activity or civil disturbances, that fall under the domain of public security²⁷⁰. In some cases, however public security effectively overlaps with national security, such as in relation to particularly serious crimes, such as terrorism and organized crime²⁷¹. This creates a functional overlap where certain national security-related measures (e.g., surveillance strategies, data retention) might fall under both public security and national security, leading to interpretative complexities and legal uncertainties²⁷². Moreover, the distinction between national and public security is becoming

²⁶⁶ Among others, see Directive 2002/58/EC, Article 1(3); Council Framework Decision 2008/977/JHA, Recital 5. See also Recital 44, 62 and Articles 13, 15, 16 of the LED and Article 72 TFEU.

²⁶⁷ TEU, Article 4(2).

²⁶⁸ CJEU, *La Quadrature du Net*, cited, paragraph 135.

²⁶⁹ A threat to national security is distinguishable by one to public security “*by its nature, its seriousness, and the specific nature of the circumstances of which it is constituted, from the general and permanent risk of the occurrence of tensions or disturbances, even of a serious nature, that affect public security, or from that of serious offences being committed*”. CJEU, *La Quadrature du Net*, cited, paragraph 136-137.

²⁷⁰ CJEU, Judgement 5 April 2022, *G.D. v The Commissioner of the Garda Síochána and Others*, Case C-140/20, ECLI:EU:C:2022:258, paragraph 62: “*It should also be observed that, unlike crime, even particularly serious crime, a threat to national security must be genuine and present, or, at the very least, foreseeable, which presupposes that sufficiently concrete circumstances have arisen to be able to justify a generalised and indiscriminate measure of retention of traffic and location data for a limited period of time. Such a threat is therefore distinguishable, by its nature, its seriousness, and the specific nature of the circumstances of which it is constituted, from the general and permanent risk of the occurrence of tensions or disturbances, even of a serious nature, that affect public security, or from that of serious criminal offences being committed*”.

²⁷¹ CJEU, Judgement of 21 December 2016, *Tele2 Sverige AB v Post-och telestyrelsen and Secretary of State for the Home Department v Tom Watson and Others*, Joined Cases C-203/15 and C-698/15, ECLI:EU:C:2016:970, preliminary rulings on whether national legislation imposing obligations on providers of electronic communications services to retain the traffic and location data of all users for a certain period is compatible with EU law. Some authors stressed that the definition of public security presented in this ruling is quite broad, encompassing national security as well. C. KUNER, F. CATE, O. LYNSKEY, C. MILLARD, N. NI LOIDEAIN, D. SVANTESSON, *An Unstoppable Force and an Immoveable Object? EU Data Protection Law and National Security*, International Data Privacy Law, 8(2), 2018, 151-167. Available at: <https://www.repository.law.indiana.edu/facpub/2692>

²⁷² Given the increasingly transnational character of crime, which poses threats both to the security of European citizens and to the Union's fundamental economic interests, criminal matters have progressively become the subject of EU policy and regulation. Initially, the European Union introduced a specific legal framework for Justice and Home Affairs (JHA)

increasingly blurred at the institutional level, as national security agencies are collaborating ever more closely with law enforcement authorities²⁷³. Such overlap becomes problematic when legislative acts apply to matters of public security but do not, or may not, apply where national security is concerned. As anticipated, the AI Act, in line with the LED, explicitly leaves the possibility to Member States to exclude national security from the scope of application of its provisions²⁷⁴. Therefore, if a national security agency uses real-time RBI systems for national security purposes, such as gathering intelligence, this would be excluded from the scope of the AI Act, and the use of such systems may potentially be permitted within Member States²⁷⁵.

The European Commission acknowledged the difficulty in differentiating between the two concepts in its *Guidelines on Prohibited Artificial Intelligence Practices*, emphasizing that “*the clear delineation of the national security exclusion is particularly important where AI systems are placed on the market, put into service or used for law enforcement purposes that fall within the scope of the AI Act*”²⁷⁶. In other words, while the use of AI systems for national security purposes falls outside the scope of the AI Act, the same does not apply to AI systems used for the protection of public security which falls within the broader notion of law enforcement, a sector that is covered by the Regulation. Therefore, the purpose of the AI system needs to be considered in order to determine whether it falls within the remit of the AI Act, rather than the authority carrying out the activities. Clarifying the differences between the two concepts is particularly important in relation to the prohibitions established in Article 5(1)(d) and (h) of the AI Act, which concern, respectively, individual crime predictions and assessments, and the use of real-time remote biometric identification systems for law enforcement

under the so-called “third pillar.” This was later developed into the Area of Freedom, Security and Justice (AFSJ), which relies primarily on cooperation among the Member States in criminal matters. With the entry into force of the Lisbon Treaty, the former pillar structure was abolished, and a dedicated Title on the AFSJ was established in the Treaties. The Lisbon reform thus both safeguarded respect for national legal orders and expanded the Union’s competences in the field of security and justice. The first coordinated European policy on security dates back to the late 1990s. The Tampere European Council (15–16 October 1999) marked a milestone, setting out the foundations of an EU approach to internal security. Following the terrorist attacks of 11 September 2001, a more comprehensive European Security Strategy was elaborated at Union level (European Security Strategy: A Secure Europe in a Better World, Council of the European Union, 2009). The Hague Programme (2005) subsequently identified priorities for the development of the AFSJ. This was followed by a series of multiannual policy frameworks, including the Stockholm Programme (2010–2014), which defined the Union’s security goals and priorities, the European Agenda on Security (2015), and the Security Union Strategy (2020–2025). Most recently, in 2020, the European Commission adopted a dedicated Counter-Terrorism Agenda, further consolidating the EU’s strategic approach to security.

²⁷³ SAJFERT, QUINTEL, *Data Protection Directive (EU) 2016/680 for Police and Criminal Justice Authorities*, cited, 4. A. ZAVRŠNIK, *Blurring the Line between Law Enforcement and Intelligence: Sharpening the Gaze of Surveillance?*, in *Journal of contemporary European research*, 9(1), 2013, 181–202. H. ADEN, *Information Sharing, Secrecy and Trust among Law Enforcement and Secret Service Institutions in the European Union*, in *West European Politics*, 41(4), 2018, 981–1002.

²⁷⁴ Similarly, the Italian Law No. 132/2025 explicitly excluded national security from its scope of application (Article 6).

²⁷⁵ It will depend on the national rules on the topic.

²⁷⁶ European Commission, *Guidelines on Prohibited Artificial Intelligence Practices*, paragraph 27

purposes²⁷⁷, since it is precisely in these areas that overlaps between public security and national security are most likely to arise.

The highlighted overlap may prove problematic, as deployers could claim that an AI system is used for national security purposes in order to fall outside the scope of the Regulation and thereby avoid its obligations. To prevent such borderline cases, the Regulation provides that if an AI system placed on the market, put into service, or used for national security (including military or defence) purposes is subsequently employed, whether temporarily or permanently, for other purposes (so-called “dual-use systems”), such as for public security purposes, that system will fall within the scope of the AI Act. In that case, the provider and the deployer of the AI system must ensure compliance of the AI system with the Regulation²⁷⁸. It is noteworthy, however, that the introduction of this national security exclusion has been severely criticised by some authors, stating that it goes against the CJEU jurisprudence²⁷⁹ and that it opens possible abuses by the Member States²⁸⁰.

2. Delimiting the concept of “law enforcement authorities” under the European Union Framework

The AI Act defines law enforcement authorities reproducing the wording of Article 3(7) of the LED, namely “*any public authority competent for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the*

²⁷⁷ *Ibidem*.

²⁷⁸ AI Act, Recital 24.

²⁷⁹ A. GIANNINI, S. TAS, *AI Act and the Prohibition of Real-Time Biometric Identification. Much ado about nothing?*, 10 December 2024. Available at: <https://verfassungsblog.de/ai-act-and-the-prohibition-of-real-time-biometric-identification/>.

²⁸⁰ This exclusion has been the subject of substantial criticism in the literature, as it risks undermining the protection of individuals precisely in one of the most intrusive domains of AI deployment. Scholars have highlighted that the AI Act, in its entirety, does not apply to systems deployed for military purposes or national security. As Nathalie A. Smuha and Karen Yeung observe, this carve-out leaves a critical gap in the regulatory framework. N.A. SMUHA, K. YEUNG, *The European Union’s AI Act Beyond Motherhood and Apple Pie?*, in N.A. SMUHA (Ed.), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*, Cambridge University Press, 2025, 228-258. C. RUDSCHIES, I. SCHNEIDER, *The Long and Winding Road to Bans for Artificial Intelligence: From Public Pressure and Regulatory Initiatives to the EU AI Act*, in *Digital Society*, 4, 2025, Article 57. The authors note that civil society organisations have characterised this “blanket exemption” as a “significant loophole”, potentially enabling practices such as mass surveillance to be justified under the guise of safeguarding national security. Recent commentary has echoed these concerns, stressing that the exemption leaves serious human rights vulnerabilities within the AI Act L.L. CABRERA, *EU’s much-heralded AI Act agreed by EU Parliament - but serious human rights holes in law remain*, Euractiv, 2024. Critics on the exclusion have been heavily criticised by European Center for Not-for-Profit Law, *Packed with loopholes: why the AI Act fails to protect civic space and the rule of law*, 3 April 2024. Available at: <https://ecnl.org/news/packed-loopholes-why-ai-act-fails-protect-civic-space-and-rule-law>. European Disability Forum, *EU’s AI Act fails to set gold standard for human rights*, 3 April 2024. Available at: <https://www.edf-feph.org/publications/eus-ai-act-fails-to-set-gold-standard-for-human-rights/#:~:text=While%20national%20security%20can%20be,largely%20digital%20rights%2Dfree%20zone>. L.L. CABRERA, *EU AI Act Brief – Pt. 2, Privacy & Surveillance*, 30 April 2024, Available at: <https://cdt.org/insights/eu-ai-act-brief-pt-2-privacy-surveillance/>.

prevention of threats to public security” (i.e. law enforcement purposes) or any other body or entity entrusted by Member State law to exercise public authority and public powers for the same purposes. Likewise, reference to data protection legislation is vital, as defining this notion is also required for the purposes of the AI Act. The definition of “*competent authority*” plays a pivotal role in the legal architecture governing data protection²⁸¹. While most Member States have incorporated the LED definitions into their national legislation, variations have emerged in how the concept of “*competent authorities*” is transposed, resulting in a fragmented legal landscape. As a result, law enforcement functions are often carried out by a decentralised network of actors, each potentially operating under distinct oversight mechanisms, legal frameworks, and technical capacities.

Article 3(7)(a) clearly covers traditional public authorities, namely judicial authorities (e.g. criminal courts), police forces, public prosecutors and other law enforcement authorities²⁸², such as customs officers²⁸³. For these authorities, law enforcement competencies are fundamental, meaning that their primary tasks and the essence of their functioning lie within the law enforcement field²⁸⁴. In addition to the competent authorities sub a), the LED extends its scope of application to include any other body or entity, including those of a private nature, that have been entrusted with law enforcement tasks by national legislation²⁸⁵. Therefore, the designation of competent authorities depends on the specific Member State and the authorities to which it has entrusted public authority for carrying out law enforcement tasks. In the case of Ireland, for example, “*competent authorities*” include also public transport operators (for example, Dublin Bus, Luas) when processing data relating to ticket fines/penalties and local authorities (for example, city and county councils) when processing data relating to parking or litter fines²⁸⁶.

The most challenging situations arise in relation to authorities of a hybrid nature, such as Financial Intelligence Units (FIUs)²⁸⁷, Passenger Information Units (PIUs) or other bodies involved in

²⁸¹ Under the LED only a “competent authority” may qualify as a controller. See LED, Article 3(8).

²⁸² LED, Recital 11.

²⁸³ TFEU, Article 87(1).

²⁸⁴ As stated above, it is becoming increasingly difficult to distinguish between law enforcement authorities responsible for public security and national security authorities, the latter of which are traditionally the intelligence services.

²⁸⁵ The initial proposal for the LED closely mirrored the approach set out in the Council Framework Decision 2008/977/JHA. However, compared to that earlier instrument – which the LED repealed and replaced – the Directive marks a significant advance in harmonizing data protection standards across the EU. While the Framework Decision was limited to cross-border data processing, the LED establishes a comprehensive framework that extends to both cross-border and domestic processing of personal data for criminal law enforcement purposes. In addition, the LED introduces a unified set of overarching rules. This represents a major improvement over the previous fragmented system, under which each EU legislative act in the field of criminal law enforcement contained its own, often divergent, data protection provisions.

²⁸⁶ Data Protection Commission, *What is a competent authority under the LED?*. Available at: https://www.dataprotection.ie/en/faqs/general/what-competent-authority-under-led?utm_source=chatgpt.com.

²⁸⁷ For a description of such authorities and their tasks see Committee of Experts on the Evaluation of Anti-Money Laundering Measures and the Financing of Terrorism, *Financial Intelligence Units*. Available at: <https://www.coe.int/en/web/moneyval/implementation/fiu>. See also, Banca d’Italia, *The Role of the Financial Intelligence*

combating crime as part of their broader mandate. Those include, for example, tax authorities, telecommunication operators, and airline companies²⁸⁸. During the expert group meeting on 7 November 2016, certain Member States observed that the processing of personal data by FIUs could potentially be subject to either the GDPR or the LED, contingent upon the FIU's nature and its operational capacity as an administrative or law enforcement authority²⁸⁹. Indeed, while FIUs are in certain cases established as administrative authorities, in others they are not²⁹⁰. This gives rise to uncertainty as to whether, and under which conditions, they qualify as competent authorities for the purposes of the LED²⁹¹.

Another interesting example is represented by tax authorities. In general, tax authorities are not considered law enforcement authorities since they typically carry out administrative enforcement rather than criminal law enforcement. Whether tax authorities are considered law enforcement authorities depends on whether national law designates them as carrying out law enforcement tasks, such as criminal investigations, the prosecution of tax fraud and the execution of criminal penalties or when they process data *on behalf of* a competent authority for law enforcement purposes. If so, these authorities may be considered competent authorities under the LED for (certain) processing activities²⁹². Similarly, banks may act on behalf of law enforcement authorities when they are requested

Unit (FIU). Available at: <https://uif.bancaditalia.it/sistema-antiriciclaggio/uif-italia/index.html?com.dotmarketing.htmlpage.language=1>.

²⁸⁸ As far as airline companies are concerned, “*The GDPR applies to the airline companies as the passenger name record (‘PNR’) data are collected for commercial purposes and further processed in order to comply with a legal obligation to which the airline companies are subject. Under Directive 2016/681 on the use of PNR data for the prevention, detection, investigation and prosecution of terrorist offences and serious crime, data processed by private entities for their own commercial purposes are then transferred (‘push’) to the authority requesting them; it is the processing of PNR data by the passenger information units to whom the data are transferred, and by competent authorities, which is then subjected to a standard of protection of personal data under national law in line with CFD 2008/977/JHA (or legislation currently in force or that will replace it, and hence in the near future the specific data protection requirements laid down in Directive 2016/680)*”. CARUANA, *The reform of the EU data protection framework in the context of the police and criminal justice sector*. As the author emphasizes, this delineation of scope should likewise extend to the processing and retention of data by telecommunications operators and financial institutions.

²⁸⁹ “[I]n some Member States the FIUs set up under the Anti-Money Laundering Directive are administrative authorities, while in others they are considered as law enforcement authorities”. Minutes of the third meeting of the Commission expert group on the Regulation (EU) 2016/679 and Directive (EU) 2016/680. T. QUINTEL, *Data protection rules applicable to Financial Intelligence Units: still no clarity in sight*, in *ERA Forum*, 23, 2022, 53–74. Available at: <https://doi.org/10.1007/s12027-021-00697-z>. T. QUINTEL, *Removing Obstacles for Cooperation between EU Financial Intelligence Units: The Data Protection Directive for Police and Criminal Justice Authorities as fitting Data Protection Instrument*, May 2019. Available at SSRN: <https://ssrn.com/abstract=3855507> or <http://dx.doi.org/10.2139/ssrn.3855507>.

²⁹⁰ VOGIATZOGLOU, MARQUENIE, *Assessment of the implementation of the Law Enforcement Directive*, cited, 22. For instance, Belgium and Spain include FIUs among the competent authorities for the purposes of the LED because they are public authorities responsible for processing personal data relating to crimes covered by the Anti-Money Laundering (AML) Directive. By contrast, national FIUs in other Member States are not considered competent authorities for the purposes of the LED, and are therefore only subject to the GDPR.

²⁹¹ M. BREWCZYŃSKA, *Financial Intelligence Units: Reflections on the Applicable Data Protection Legal Framework*, in *Computer Law & Security Review*, 43, 2021, Article 105612.

²⁹² By contrast, in relation to tax authorities, in *SIA SS v Valsts ieņēmumu dienests* ruling, the Court – following the considerations of Advocate General Bobek – held that tax authorities “when they request an economic operator to disclose to them personal data relating to certain taxpayers for the purposes of the collection of tax and combating tax fraud” they

to conduct certain actions to counter certain crimes in specific cases under the instructions and supervision of law enforcement authorities²⁹³.

The inclusion of private entities within the scope of the definition reflects a broader trend across many European countries, where law enforcement functions are increasingly being delegated or outsourced to private actors. As a result, the LED's expanded scope aims to ensure that the fundamental rights to data protection are preserved, even when traditionally public law enforcement functions are carried out by private organizations operating under legal mandates. In order for a private entity or body to qualify as a competent authority under Article 3(7) of the LED, it must be formally entrusted by law to exercise public authority and public powers in the context of law enforcement. This means the entity must not merely act in cooperation with public authorities but must itself be legally mandated to perform tasks that involve the exercise of state-like coercive powers or decision-making functions typical of public institutions. This is, for example, the case of security service contractors for sporting events²⁹⁴. Consequently, private entities that are legally obliged to collect or process personal data only at the request of public authorities do not fall within the definition of competent authorities.

As clarified by the European Commission, competent authorities are “*private bodies, on which the law confers special powers beyond those which result from the normal rules applicable in relations between individuals and/or by the possibility of exercising the power of coercion*”²⁹⁵. This means that the definition extends beyond traditional law enforcement authorities to cover any bodies or entities, whether public or private, that perform public functions identified in Article 3(7) LED. The scope may also encompass various other competent authorities that are entrusted by national laws to exercise these functions, or that process data on behalf of a competent authority for law enforcement purposes, such as bank institutions when performing their tasks under anti-money laundering obligations²⁹⁶.

During the proposal phase, both the European Data Protection Supervisor and the Article 29 Working Party expressed concerns about broadening the definition of competent authorities to include entities

may not “*be regarded as a ‘competent authority’ within the meaning of Article 3(7) of Directive 2016/680*”. CJEU, Judgement of 24 February 2022, *SIA SS v Valsts iepēmumu dienests*, Case C-175/20, ECLI:EU:C:2022:124, paragraph 44.

²⁹³ *Guidelines on Prohibited Artificial Intelligence Practices*, cited, paragraph 324.

²⁹⁴ CARUANA, *The reform of the EU data protection framework in the context of the police and criminal justice sector*, cited, 5. CNIL, “*Law Enforcement Directive*”: *What Are We Talking About?*, cited.

²⁹⁵ European Commission, *Communication from the Commission to the European Parliament and the Council, First report on application and functioning of the Data Protection Law Enforcement Directive (EU) 2016/680 (‘LED’)*, COM(2022) 364 final, 25 July 2022, 10. By contrast, in its interpretation of the Passenger Name Record (PNR) Directive, the CJEU excluded private entities from the scope of the LED, reasoning that they are “*neither in charge of exercising public authority nor entrusted with public powers by that directive*”. This interpretation suggests that, in fields governed by EU law, it is the EU legislature – not the Member States when transposing the instrument – that must confer public powers or authority on a private body for it to fall within the definition of a competent authority under Article 3(7)(b) LED.

²⁹⁶ On the topic of anti-money laundering, see European Central Bank, *Anti-money laundering*. Available at: <https://www.bankingsupervision.europa.eu/framework/anti-moneylaundering/html/index.en.html>; European Commission, *Anti-money laundering and countering the financing of terrorism at EU level*. Available at: https://finance.ec.europa.eu/financial-crime/anti-money-laundering-and-countering-financing-terrorism-eu-level_en.

not strictly involved in criminal matters, which would otherwise fall under the scope of the GDPR²⁹⁷. Since the LED grants to Member States a broader margin of discretion and is, in some respects, less stringent than the GDPR, any expansion of its scope of application may pose challenges to the effective protection of personal data²⁹⁸.

Despite some interventions of the CJEU, considerable uncertainty and confusion remain as to which authorities fall within the definition in Article 3(7) LED²⁹⁹. The identification of national competent authorities is indeed largely contingent on definitions of concepts such as “criminal offence”, “public security”, “public authority” and “public power”, which are strictly dependent on the national legal order. As far as the LED is concerned, different understandings of these concepts may lead to different types of national authorities falling within the scope of the Directive³⁰⁰. Every Member State has opted for different approaches, some of them explicitly listed competent authorities within national legislation, some others opted for broader or restrictive definitions of competent authorities³⁰¹. Slovenia, for example, explicitly lists national competent authorities, alongside an open provision that covers any authority with jurisdiction over crimes and criminal penalties³⁰². In Germany, the implementation of the LED occurred both at the federal and Länder level, with horizontal acts³⁰³ and

²⁹⁷ European Data Protection Supervisor, *Opinion 6/2015*, cited, 6. “[T]he notion of “competent authority”, as defined under Article 3(13) [actual Article 3(7)] should also remain as limited as possible: the performance of law enforcement tasks by non-public entities and organisations should be subject to the Regulation and not the Directive” since they “do not need a specific regime” as “the main and original purpose of the collection of these data is totally different from the prevention, investigation, detection or prosecution of criminal offence”. Therefore, the processing of personal data under the LED should be “limited to the areas where specific rules are really necessary”.

²⁹⁸ VOGIATZOGLU, MARQUENIE, *Assessment of the implementation of the Law Enforcement Directive*, cited, 22.

²⁹⁹ European Data Protection Board, *Contribution of the EDPB to the European Commission’s evaluation of the Data Protection Law Enforcement Directive (LED) under Article 62*, 14 December 2021, 22. Some authors proposed a “publicly available centralised EU registry of competent authorities” which would bring more clarity with respect to the implementation of this provision across the EU. E. KOSTA, R. LEENES, I. KAMARA (Ed.), *Research handbook on EU data protection law*, Edward Elgar Publishing, 2022, 108.

³⁰⁰ While the GDPR is a regulation, it has general application and it is binding in its entirety and directly applicable in all Member States, leaving less room for discretion, the LED is a directive and it is binding as to the result it sets out, which can be achieved in a form and with methods chosen by each Member State. In addition, the LED’s proximity to the criminal justice process, which is characterised by a variety of different national regimes, has led to many differences between Member States. For an in-depth comparative examination of police data processing in five countries see *De verwerking van politiegegevens in vijf Europese landen*, Groningen, 20 november 2020. Available at: <https://repository.wodc.nl/bitstream/handle/20.500.12832/3025/3031-de-verwerking-van-politiegegevens-in-vijf-Europese-landen-volledige-tekst.pdf?sequence=7&isAllowed=y>.

³⁰¹ For a detailed analysis of national legislations implementing the LED see *Study on the transposition of Directive (EU) 2016/680*, Tipik Legal and Milieu Consulting, March 2023. Available at: https://commission.europa.eu/publications/publications-law-enforcement-directive-led_en. For national transposition measures communicated by Member States see the official page EUR-Lex. Available at : <https://eur-lex.europa.eu/legal-content/en/NIM/?uri=celex:32016L0680>.

³⁰² As far as courts are concerned, only some provisions of the LED are considered applicable when they process personal data in criminal matters.

³⁰³ An example is the Federal Data Protection Act of 30 June 2017 (BDSG), in Part 3 §45(1) it generally refers to “public bodies competent for the prevention, investigation, detection or prosecution of criminal or administrative offences or the execution of criminal or administrative penalties, as far as they process data for the purpose of carrying out these tasks”. In this definition, the Federal Criminal Police Office (Bundeskriminalamt, “BKA”), the Customs Investigation Service (Zollfahndungsamt, “ZFA”), and other federal law enforcement bodies”.

sectoral laws that apply only to specific competent authorities³⁰⁴. In the United Kingdom, in the Data Protection Act – which included both a full implementation of the GDPR and the LED – the competent authorities falling under this definition from a public-sector perspective are explicitly enlisted in Schedule 7 of the Act, which provided a long series of governmental bodies that met the requirements of the LED³⁰⁵. Similarly, the Belgian *Loi du 30 juillet 2018 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel* presents a detailed definition of competent authority within the meaning of the LED through a seemingly exhaustive list of entities³⁰⁶. Conversely, Italy did not list competent authorities by name in the transposing legislation, relying on the functional definition mirroring Article 3(7) of the LED³⁰⁷. The definition covers the Italian police forces (Polizia di Stato, Carabinieri, Guardia di Finanza), judiciary (courts and public prosecutors when acting in criminal matters), penitentiary authorities, and any other body given law-enforcement powers by Italian law.

In any case, not all public bodies engaged in police-related activities should automatically be considered competent authorities under the LED. In particular, entities engaged in administrative police functions – such as those involved in authorising certain activities, imposing related sanctions, or overseeing vehicle circulation – do not always fall within the scope of the derogatory regime established by the Directive. In the Italian context, these conclusions align with the Constitutional Court's case law, which draws a distinction between police activities aimed at maintaining public order and security, on the one hand, and administrative policing, on the other. The Court has emphasised that the maintenance of public order and security relates to the prevention of crime and the safeguarding of fundamental legal interests that underpin peaceful coexistence within society³⁰⁸. In contrast, administrative police functions concern preventive measures aimed at avoiding harm or damage to individuals or their properties, provided that such measures do not involve interests specifically protected for reasons of public order or security³⁰⁹.

The Data Protection Commission (DPC), the Irish Data Protection Authority, intervened on the scope of application of the LED. For the Irish DPA, the concept of competent authorities “*is not limited to*

³⁰⁴ In Bremen, Hamburg, and Saarland, no horizontal legislation can be identified. In Baden-Württemberg, by contrast, the main instrument implementing the LED is the Act on the Protection of Personal Data for Processing by the Judicial Authorities of the Land (Landesdatenschutzgesetz für Justiz- und Bußgeldbehörden – LDSG-JB). However, its scope is limited to the competent authorities falling under the responsibility of the Ministry of Justice. Consequently, the LDSG-JB does not extend to the processing of personal data by the police, which instead falls under the Police Act (Polizeigesetz – PolG (BW)).

³⁰⁵ References of governmental bodies can be amended.

³⁰⁶ Belgian *Loi du 30 juillet 2018 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel*, Article 26(7).

³⁰⁷ Legislative Decree No 51/2018, Article 2(1) lett. g).

³⁰⁸ *Ex multis* Corte Costituzionale, Judgement No. 290, 2001; Corte Costituzionale Judgement No. 129, 2009.

³⁰⁹ *Ex multis*, Corte Costituzionale, Judgement No. 218, 1988; Corte Costituzionale, Judgement No. 129, 2009.

*processing by bodies typically considered 'law enforcement authorities' (such as An Garda Síochána), but extends "to any processing for law enforcement purposes carried out by a public or private body that fits the definition of 'competent authority' (such as local authorities when prosecuting litter fines, or Dublin Bus in relation to ticket offences)"*³¹⁰. This means that a potentially very large number and variety of bodies might fall under the scope of the regime, which will need to be assessed on a case-by-case basis.

While most of the Member States' laws comprehensively cover any competent authority processing of data for LED purposes, mirroring the definition enshrined in Article 3(7) LED, some Member States have chosen to narrow the scope of application of the LED in their national transposing measures. In some instances, it has been done by specifying certain types of data to which the LED applies³¹¹ or by explicitly excluding relevant competent authorities from the national provisions implementing the LED³¹². For example, some Member States, such as Bulgaria and Greece³¹³, have chosen to limit the application of the LED to public authorities, thereby excluding private entities from its scope.

Therefore, in certain cases the classification is uncontroversial, such as with police forces or criminal justice bodies, including public prosecutors and criminal courts. However, in other instances the delineation proves far less straightforward. This difficulty arises, for example, with regard to tax authorities or FIUs, whose activities may intersect with law enforcement functions without being primarily oriented towards them. Comparable issues emerge in relation to private entities – such as telecommunication operators, or airline companies – raising the question of whether, and under what conditions, such actors may fall within the scope of “law enforcement authorities” for the purposes of the LED and, by extension, the AI Act.

Unlike the LED, the AI Act provides some useful interpretative guidance with respect to these notions. As regards public authorities, Recital 59 partially addresses the uncertainties left unresolved by the LED. Specifically, it clarifies that *“AI systems specifically intended to be used for administrative proceedings by tax and customs authorities as well as by financial intelligence units carrying out administrative tasks analysing information pursuant to Union anti-money laundering law should not be classified as high-risk AI systems used by law enforcement authorities for the purpose of prevention, detection, investigation and prosecution of criminal offences”*. From this formulation, it may be

³¹⁰ Data Protection Commission, *What is a competent authority under the LED?*, cited.

³¹¹ In the Netherlands, national legislation refers only to specific categories of personal data, such as ‘police data’ (i.e., personal data processed in the context of police tasks), as well as ‘judicial,’ ‘criminal,’ ‘judicial criminal,’ or ‘enforcement’ data. In Slovenia, the main national implementing act excludes the application of several of its provisions with respect to courts acting in criminal matters. *Study on the transposition of Directive (EU) 2016/680*, cited, 8.

³¹² In the Czech Republic, the Municipal Police fall outside the scope of the transposing legislation. In Slovenia, not all provisions of the LED apply when courts act in criminal matters. *Ibidem*.

³¹³ In relation to such aspect, the Greek supervisory authority opined that this approach goes against the wording of the LED. VOGIATZOGLU, MARQUENIE, *Assessment of the implementation of the Law Enforcement Directive*, cited, 21.

inferred that when such authorities perform tasks directly connected to criminal activities, they are to be considered law enforcement authorities within the meaning of the AI Act. Conversely, when acting exclusively in their administrative capacity, they fall outside this definition.

With regard to private entities, Article 3(46) AI Act explicitly provides that law enforcement activities may be carried out either by “*law enforcement authorities*” or “*on their behalf*”. Further clarification is offered by the *Guidelines on Prohibited Artificial Intelligence Practices*, which stress that, under the AI Act, entities, bodies, or persons other than traditional law enforcement authorities may perform law enforcement activities where Member States have entrusted them with public authority and powers for the purposes listed in the Regulation³¹⁴. The notion of acting “*on behalf of*” is understood to imply that national legislation may entrust the performance of a law enforcement task (or part thereof) to another entity or person, including private parties, or requests their support in specific cases. Crucially, such delegation requires that the law enforcement authority retains control by issuing instructions on all major aspects and by exercising supervision over the entrusted entity, since these conditions are inherent to the very notion of acting “*on behalf of*”.

By way of example, delegation may involve: public transport companies instructed to ensure security on transport networks under the law enforcement authority’s supervision; sports federations requested to provide security at major events under direct instruction and supervision of the law enforcement authority; or banks tasked, in specific cases, with taking measures to counter certain crimes under the supervision of competent authorities. In all such instances, the private entity acts within the scope of law enforcement purposes as it operates under the authority and supervision of a law enforcement body. Conversely, where private entities act on their own behalf, for instance, when banks or other companies independently detect and counter crimes such as fraud or money laundering in pursuit of their own compliance obligations, they cannot be regarded as exercising law enforcement functions within the meaning of the AI Act.

3. Law enforcement and human rights: striking a delicate balance

The above legal analysis is pivotal to interpreting and contextualizing Articles 3(45) and 3(46) of the AI Act, which establish key definitional and conceptual foundations for understanding the role of AI in law enforcement contexts. Building on this foundation, it becomes necessary to shift the discourse toward law enforcement from a different point of view, namely the concrete operational activities undertaken by law enforcement authorities (hereinafter “LEAs”) and the manner in which these actions

³¹⁴ *Guidelines on Prohibited Artificial Intelligence Practices*, cited, paragraphs 322-325.

directly or indirectly affect the enjoyment and protection of fundamental rights by individuals. To fulfil their responsibilities and safeguard the constitutional interests of public order and security, LEAs are vested with extensive powers, including coercive measures and the use of force. Yet the exercise of such powers often entails limitations or intrusions upon individuals' fundamental rights³¹⁵. In recent decades, advances in the technological process have significantly expanded the powers available to LEAs in carrying out their duties. Alongside these developments, their capacity to intrude into individuals' private lives and to interfere with fundamental human rights has also grown.

In the last decades, scholars have extensively examined the transformation of public security strategies, particularly focusing on the transition from *reactive* policing to *pre-emptive* approaches³¹⁶. Lucia Zedner conceptualises this as a transition from a "post-crime" to a "pre-crime" society, one in which the anticipation and mitigation of potential risks increasingly take precedence over addressing offences already committed. In the post-crime model, the focus is on crimes, offenders, victims, investigation, trial, and punishment. By contrast, the pre-crime paradigm shifts the temporal lens forward, seeking to anticipate and prevent events that have not yet occurred, and may never occur. This preventive logic is characterised by "*calculation, risk and uncertainty, surveillance, precaution, prudentialism, moral hazard, prevention*"³¹⁷, all underpinned by an overarching commitment to security as the primary objective.

The fact that security issues are increasingly framed as risks, determines that they can be anticipated and mitigated by preventive measures, the ones that are based on information-gathering and that

³¹⁵ On this topic see Council of Europe, *The European Convention On Human Rights and Policing*, December 2013. Available at: https://www.echr.coe.int/documents/d/echr/handbook_european_convention_police_eng. An interference with human rights can be seen also in situations which do not arrive at unwarranted arrests, prolonged detention, and misuse of search powers.

³¹⁶ The policing literature on the topic is huge. C.S. STEIKER, *Foreword: The Limits of the Preventive State*, in *The Journal of Criminal Law and Criminology*, 88(3), 1998, 771-808. D. GARLAND, *The Culture of Control: Crime and Social Order in Contemporary Society*, University of Chicago Press, 2002; B. HARCOURT, *The Shaping of Chance: Actuarial Models and Criminal Profiling at the Turn of the Twenty-First Century* in *The University of Chicago Law Review*, 70(1), 2003, 105-128; I. BROWN, D. KORFF, *Privacy & Law Enforcement*, February 2004. Available at SSRN: <https://ssrn.com/abstract=3737428> or <http://dx.doi.org/10.2139/ssrn.3737428>. L. ZEDNER, *Pre-Crime and Post Criminology?*, in *Theoretical Criminology*, 11(2), 2007, 261-281; R. VAN BRAKEL, P. DE HERT, *Policing, surveillance and law in a pre-crime society: Understanding the consequences of technology based strategies*, *Cahiers Politicestudies*, 3(20), 2011, 163-192; A. ASHWORTH, L. ZEDNER, *Prevention and Criminalization*, in *New Criminal Law Review: An International and Interdisciplinary Journal*, 15(4), 2012, 542-571; I. KERR, J. EARLE, *Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy*, in *Stanford Law Review*, 66, 2013, 65-72. J. OGG, *Preventive Justice and the Power of Policy Transfer*, Springer, 2015; J. MCCULLOCH, D. WILSON, *Pre-crime Pre-emption, precaution and the future*, Routledge, 2015; V. MITSILEGAS, *The Transformation of Privacy in an Era of Pre-emptive Surveillance*, 20(1), 2015, 35-57. R. VAN BRAKEL, *Pre-Emptive Big Data Surveillance and its (Dis)Empowering Consequences: The Case of Predictive Policing*, in B. VAN DER SLOOT ET AL (Ed.), *Exploring the Boundaries of Big Data*, Amsterdam University Press, 2016, 117-141. R. SELMINI, *Sicurezza urbana e prevenzione della criminalità in Europa: alcune riflessioni comparate*, in *Polis, Ricerche e studi su società e politica*, 1, 1999, 69-76.

³¹⁷ ZEDNER, *Pre-Crime and Post Criminology?*, cited, 262

operate prior to any formal suspicion of a concrete criminal offence³¹⁸. This “*preventive-security governance*”³¹⁹ is becoming more pervasive through omnipresent surveillance mechanisms aimed at managing risks, significantly impacting core fundamental rights and freedoms by normalising bulk data collection, prolonged retention, and cross-database interoperability, thereby shifting coercive power to the pre-suspect stage.

Law enforcement is often characterised as a data-intensive activity, given that LEAs during their routine operations collect and process vast quantities of data. This includes data gathered during investigations, from LEA case files, reports on their daily activities, and calls for service. Such repositories constitute a rich dataset, offering significant potential for the development of data-driven tools aimed at enhancing the efficiency and overall effectiveness of policing activities. Despite the advantages of using data-driven tools, it is generally acknowledged that the police’s use of data is socially sensitive, due to the delicate relationship existing between LEAs and citizens or individuals in general, which is inherently characterised by an imbalance of power. Police officers and other LEAs operate with state-granted authority, which includes coercive powers such as detention, search, seizure, and, in some cases, the use of force, which private individuals do not possess. Because of this imbalance, the law – especially constitutional provisions and criminal procedural law – imposes safeguards to prevent abuses and ensure that the exercise of police powers remains reasonable, proportionate, necessary and consistent with human rights norms.

Integrating AI into policing has taken law enforcement to a new level, enabling a faster and deeper analysis of data to identify trends, patterns and correlations, and extrapolate valuable information that LEAs can use to carry out police operations more effectively and efficiently. While AI can give law enforcement unprecedented reach, speed, and analytical power, individuals are left with the same, if not fewer, tools to understand, resist, or hold authorities accountable, deepening a structural imbalance in favour of State power. The European legislator explicitly recognises this risk, noting in the AI Act that “[g]iven their role and responsibility, actions by law enforcement authorities involving certain uses of AI systems are characterised by a significant degree of power imbalance and may lead to surveillance, arrest or [the] deprivation of a natural person’s liberty as well as other adverse impacts on fundamental rights guaranteed in the Charter”³²⁰. Building on this recognition, the AI Act

³¹⁸ P. VOGIATZOGLU, *The Era Of Pre-Crime: How Mass Data Surveillance And Predictive Policing Intersect And Interfere With Privacy, Data Protection And Due Process Rights In The EU*, Doctoral Thesis, KU Leuven Faculty of Law and Criminology, 2023, 109.

³¹⁹ M. DEN BOER, *Juggling the Balance between Preventive Security and Human Rights in Europe*, Security and Human Rights Monitor, 2. DOI: 10.1163/18750230-02602009.

³²⁰ AI Act, Recital 59.

establishes a complex and stringent regulatory framework governing the deployment of AI systems in the context of law enforcement³²¹.

The integration of AI and the preventive measures outlined above fundamentally challenge traditional rule of law and criminal justice principles³²². Contemporary policing now extends well beyond discrete, transactional encounters between officers and citizens, evolving into surveillance infrastructures that enable pervasive – or, as Christopher Slobogin terms them, “dragnet”³²³ – forms of monitoring, where data collection and analysis occur continuously, systematically, even in the absence of an individualised suspicion. As Daphna Renan notes, such policing is programmatic and ongoing, exhibiting characteristics markedly different from conventional law enforcement activities³²⁴. Nevertheless, these surveillance regimes must adhere to principles of rationality, be grounded in empirical evidence, and operate within clearly defined legislative frameworks; their implementation should be impartial, subject to public oversight to the extent compatible with effective law enforcement, and open to judicial review. Renan frames these challenges as problems of governance, contending that administrative law offers a robust framework for regulating technology-driven policing, particularly in programmatic or “panvasive”³²⁵ surveillance contexts³²⁶.

These concerns, widely discussed in American scholarship, align with the conclusions of the European Parliament, which in its Resolution of 6 October 2021 observed that traditional criminal procedural law no longer provides an adequate framework for addressing contemporary surveillance practices³²⁷. In particular, such practices now involve the large-scale collection and retention of data relating to

³²¹ I will further discuss this topic in Chapter 3.

³²² For example, the principle of presumption of innocence. A. GALETTA, *The changing nature of the presumption of innocence in today's surveillance societies: rewrite human rights or regulate the use of surveillance technologies?*, in *European Journal of Law and Technology*, 4(2), 2013. The article explores how the presumption of innocence, traditionally guaranteed only within formal criminal proceedings, is increasingly challenged by the rise of surveillance technologies. These technologies often operate preventively, subjecting individuals to monitoring, profiling, and even suspicion before any formal accusation or trial, which blurs the line between prevention and prosecution. This shift risks eroding the meaning of the presumption of innocence by normalizing scrutiny without charges. The author argues that two responses are possible: either human rights law must be reinterpreted or rewritten to extend protections to pre-charge phases, or surveillance technologies must be more strictly regulated to ensure they respect fundamental rights. Ultimately, the article suggests that a combination of regulation and legal adaptation is necessary to preserve the relevance of the presumption of innocence in surveillance societies.

³²³ C. SLOBOGIN, *Government Dragnets*, in *Law and Contemporary Problems*, 73(3), 2010, 101. Available at SSRN: <https://ssrn.com/abstract=1640108>

³²⁴ D. RENAN, *The Fourth Amendment as Administrative Governance*, in *Stanford Law Review*, 68, 2016, 1039-1129.

³²⁵ C. SLOBOGIN, *Panvasive Surveillance, Political Process Theory and the Nondelegation Doctrine*, in *Georgetown Law Journal*, 2014, 1721–1776.

³²⁶ C. SLOBOGIN, *Administrative Regulation of Programmatic Policing: Why "Leaders of a Beautiful Struggle" is Both Right and Wrong*, in *Administrative Law Review*, 75(3), 2023, 565-586.

³²⁷ European Parliament, *Resolution of 6 October 2021 on Artificial Intelligence in criminal law and its use by law enforcement authorities in criminal matters*, 2020/2016(INI), paragraph Q: “Whereas modern criminal law is based on the idea that authorities react to an offence after it has been committed, without assuming that all people are dangerous and need to be constantly monitored in order to prevent potential wrongdoing; whereas AI-based surveillance techniques deeply challenge this approach [...]”.

hundreds of thousands of individuals, even before any criminal offence has been detected or legal proceedings initiated. Since there is not even an individualised suspicion at the time of collection, the tools provided for by criminal procedural law (such as an individual-specific warrant) are insufficient to adequately protect individuals³²⁸. More precisely, criminal procedural law retains its central importance in the post-offence phase, governing the investigation and prosecution of criminal offences once a crime has been committed. However, in contexts where surveillance programmes are deployed to gather information *prior* to the emergence of any individualized suspicion, the safeguards and procedural requirements of criminal procedural law are not engaged, leaving such activities to be assessed under alternative legal frameworks³²⁹. While embedding AI-enabled policing within an administrative governance framework makes it possible to balance operational efficiency with the protection of fundamental rights, administrative law principles themselves must evolve to address the novel challenges posed by the use of AI by public authorities³³⁰. The integration of AI into public authorities' decision-making is already at the forefront of political, legal, and academic debate³³¹.

First and foremost, the high interests at stake and the potential consequences for individuals affected by mistakes caused by AI-based policing tools call to the forefront the respect for the rule of law³³². The principle of legality remains the foundation, since it requires that police actions operate strictly under, and within, the law, both substantively, by ensuring that the legal conditions for exercising such powers are met, and procedurally, by respecting the legal procedures governing the exercise of such powers. In Chapter 3, I will examine which AI systems developed for and deployed within the law enforcement sector are prohibited under European law, as well as those classified as high-risk and therefore permissible, subject to strict compliance with the conditions set forth in the Regulation.

Since the law cannot foresee every possible scenario in advance, legislators inevitably leave certain areas where the police must exercise discretion, which has been generally understood as the authority to choose among multiple lawful alternatives, granted to public authorities when it is necessary to

³²⁸ As I will further analyse in Chapter 2, the adoption of AI raises substantial legal and human rights concerns: AI's "black box" nature can obscure the reasoning behind decisions, complicating judicial review and public oversight; algorithms may replicate systemic biases embedded in historical data, disproportionately affecting certain individuals or communities; heavy reliance on automated tools risks sidelining the nuanced balancing of interests that human discretion provides.

³²⁹ C. SLOBOGIN, *Administrative Regulation of Programmatic Policing*, cited.

³³⁰ In these terms, see L. TORCHIA, *Lo Stato digitale. Una introduzione*, il Mulino, 2025, 110.

³³¹ B. MARCHETTI, *Amministrazione digitale*, in *Enciclopedia del Diritto: funzioni amministrative*, 2022, 75-109; J.B. AUBY, *Il diritto amministrativo di fronte alle sfide digitali*, in *Le istituzioni del federalismo*, 3, 2019, 619-641; I. MARTÍN DELGADO, *Automazione, intelligenza artificiale e pubblica amministrazione: vecchie categorie concettuali per nuovi problemi?*, in *Le istituzioni del federalismo*, 3, 2019, 643-66; G. ORSONI, E. D'ORLANDO, *Nuove prospettive dell'amministrazione digitale: Open Data e algoritmi*, in *Le istituzioni del federalismo*, 3, 2019, 593-617; D.U. GALETTA, J.G. CORVALÁN, *Intelligenza artificiale per un'amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *Federalismi.it*, 2019, 1.

³³² In general, the rule of law must be respected when public administrations use AI. Cfr. B. MARCHETTI, *Artificial Intelligence and Public Authorities: Does the European AI Act Protect Public Values?*, in *European Review of Public Law*, 36(1), 2024, 67-90.

balance competing public and private interests in order to reach a decision³³³. Where the law does not regulate the exercise of administrative power in exhaustive detail, public authorities have a duty to consider all relevant factors, carefully assessing and reviewing the factual and legal circumstances of each case before making a decision³³⁴.

In policing, discretionary powers are required both in determining whether the conditions for intervention are met and in selecting the most appropriate operational measures, considering the specific merits and context of each case. The integration of AI into law enforcement may transform the way discretion is exercised, narrowing the scope of individual officer discretion, as choices become increasingly influenced by algorithmic outputs³³⁵. As I will further analyse in Chapter 2, AI systems (predictive policing algorithms, face recognition, automated license plate readers, etc.) create new decision inputs. Therefore, instead of relying only on direct observation, experience, reports and call for services, the police officer might receive a hotspot-alert suggesting a higher likelihood of crime in a location, a face recognition match for someone in a crowd or an algorithmic risk score for an individual. Although the final decision still legally and ethically rests with the officer, he/she may feel compelled to act in conformity with the algorithmic output. This phenomenon, referred to as “automation bias”, entails a propensity to place undue trust in algorithmic outputs, which can result in the attenuation of human judgment, diminished capacity for error detection, the perpetuation and amplification of systemic biases, and the emergence of significant accountability deficits. Consequently, the establishment of robust procedural safeguards is essential to mitigate these risks and to guide the exercise of discretion, thereby ensuring that algorithmic tools operate as aids to – rather than substitutes for – informed human judgment.

³³³ The topic of discretion has been widely discussed in Italian administrative literature. Cfr. M.S. GIANNINI, *Il potere discrezionale della pubblica amministrazione. Concetto e problemi*, Giuffrè, 1939; For critical reflections on the that concept of discretion. cfr. F.G. SCOCA, *La discrezionalità nel pensiero di Giannini e nella dottrina successiva*, in *Rivista Trimestrale di Diritto Pubblico*, 2000, 1045; F.G. SCOCA, *Attività amministrativa*, in *Enc. Dir., Agg.*, VI, Giuffrè, 2002, 75. Cfr. also A. PIRAS, *Discrezionalità amministrativa*, in *Enc. Dir.*, XIII, Giuffrè, 1964, 65; B.G. MATTARELLA, *Discrezionalità amministrativa*, in S. CASSESE (Ed.), *Dizionario di diritto pubblico*, III, Giuffrè, 2006, 1993. An interesting reconstruction of the notion of discretion in Italian landscapes can be found in G. TERRANO, *La discrezionalità amministrativa nell'evoluzione della forma di Stato*, PhD Thesis, 2023.

³³⁴ This obligation can be summarised in the so called “duty of care” (EU General Principles).

³³⁵ On the specific topic of administrative discretion in the digital era see the interesting contribution of A. CASSATELLA, *La discrezionalità amministrativa nell'età digitale*, in *Diritto amministrativo: scritti per Franco Gaetano Scoca – I*, 2021, 675-708.

CHAPTER 2

Artificial intelligence enters law enforcement

1. Overview of the Chapter	95
2. The digitisation of police organisations.....	99
Part I: Predictive policing	105
1. The current frontier of data analytics: predictive policing.....	105
1.1. The experience of predictive policing in the United States	105
1.2. Defining predictive policing.....	109
1.2.1. Place-based predictive policing	113
1.2.2. Person-based predictive policing	116
1.3. Predictive policing in Europe	119
2. Considerations on predictive policing	126
2.1. Data and data protection concerns.....	129
2.2. Bias and non-discrimination.....	133
2.3. Errors and accuracy concerns.....	139
2.4. Transparency issues.....	140
Part II: Automated video surveillance and face recognition technology	143
1. The police use of video analytics.....	143
1.1. Image processing through Automated Video Surveillance.....	144
1.2. Identifying individuals through biometric technologies.....	151
2. An example of hard biometrics: face recognition technology	153
2.1. Face recognition technology: how does it work?	155
2.2. The use of face recognition technology: where does Europe stand?	162
3. Considerations on Automated Video Surveillance and Face Recognition Technology.....	164
3.1. Data and data protection concerns.....	164
3.2. Errors, accuracy concerns and bias	170
3.3. The ever-closer relationship between public authorities and private entities	176

1. Overview of the Chapter

Although it might seem that the discourse on the use of AI in law enforcement contexts is novel, this statement cannot be considered correct, at least by adopting a global perspective. Indeed, around the 1980s in the United States some interesting works were published on this topic. In 1986, David J. Icové wrote a short article titled *Automated Crime Profiling*¹ in which he described an AI-based computer

¹ D.J. ICOVE, *Automated Crime Profiling*, in *FBI Law Enforcement Bulletin*, 55(12), 1986, 27-30.

system used by the National Center for the Analysis of Violent Crime (NCAVC) of the Federal Bureau of Investigation (FBI). In his work, the author pointed out that the NCAVC developed, based on the insights and experience gained with the FBI's Violent Criminal Apprehension Program (VICAP) and the Arson Information Management System (AIMS), a knowledge-based expert system to assist members of the NCAVC in tracking and predicting violent crimes. Almost forty years ago he thus described a pioneering use of AI in crime analysis and criminal personality profiling. Some years later, in 1989, Edward C. Ratledge and Joan E. Jacoby published the *Handbook on Artificial Intelligence and Expert Systems in law enforcement*², where they optimistically described multiple uses of AI in law enforcement, mainly focusing on expert systems. Examples of such uses are real-time decision-making, crime-solving³, personnel management and training, program planning and design.

These works stress that the discourse on the use of AI in law enforcement is far from new, at least as far as the United States are concerned. Indeed, it was the FBI that had a fundamental role in the development of expert systems and in the use of AI-based technology in assisting human operators in carrying out activities in criminal matters. However, from the 1980s some changes can be stressed: first, advanced technologies are now also used outside the intelligence field, by local police departments; second, the amount of data available is wider; third, the algorithms used are generally more complex; fourth, the discourse around the use of AI in law enforcement is not always positive, because of expressed concerns for human fundamental rights and freedoms.

The use of AI in law enforcement is part of a wider trend across the public sector to introduce algorithmic tools to inform decision-making⁴. Some scholars have seen the use of AI in the public sector as a new phase in the process of digitising administration, laying the foundation for the transition from digital administration to algorithmic administration⁵. As I pointed out in Chapter 1, AI is not just

² E.C. RATLEDGE, J.E. JACOBY, *Handbook on Artificial Intelligence and Expert Systems in law enforcement*, Greenwood, 1989.

³ In relation to crime-solving the authors describe the application of an expert system in Baltimore County, Maryland for residential burglaries. *Ibidem*, 117-129.

⁴ The Italian literature on the topic is huge. Among others, G. AVANZINI, *Decisioni amministrative e algoritmi informatici. Predeterminazione, analisi predittiva e nuove forme di intelleggibilità*, Editoriale Scientifica, 2019. B. MARCHETTI, *Voce Amministrazione Digitale*, in *Enciclopedia del Diritto: funzioni amministrative*, Giuffrè, 2022, 75-109. R. CAVALLO PERIN, D.U. GALETTA (Eds.), *Il diritto dell'amministrazione pubblica digitale*, Giappichelli, 2020. For an interesting and detailed analysis on the uses of AI in the Italian public administration see M.B. ARMIENTO, *Pubbliche amministrazioni e intelligenza artificiale. Strumenti, principi e garanzie*, Editoriale Scientifica, 2024. For a more general perspective, see C. COGLIANESE, *Administrative Law in the Automated State*, in *Daedalus*, 150 (3), 2021, 104–120. C. COGLIANESE, D. LEHER, *Regulating by robot: Administrative Decision Making in the Machine Learning Era*, in *Georgetown Law Journal*, 105, 2017, 1147-1223. D.F. ENGSTROM, D.E. HO, C. M. SHARKEY, M.-F. CUÉLLAR, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*. Report Submitted to the Administrative Conference of the United States, February 2020. Available at: <https://www.acus.gov/document/government-algorithm-artificial-intelligence-federal-administrative-agencies>.

⁵ In the Italian context, terms such as “algorithmic administration” or “administration through algorithms” have been coined to describe the use of algorithms in the public sector. See A. SIMONCINI, *Profili costituzionali dell'amministrazione algoritmica*, in *Rivista Trimestrale di Diritto Pubblico*, 4, 2019, 1149-1188. A. DI MARTINO, *L'amministrazione per*

one thing, it is rather an umbrella term encompassing a family of very different technologies, in which different approaches can be identified in order to create machines that behave intelligently⁶. In Chapter 1 I analysed knowledge-based AI, machine learning and deep learning, as current approaches to AI. Distinguishing between different approaches to AI is crucial, since the challenges they pose to principles that govern public powers are different⁷. Moreover, it is crucial to distinguish between different uses of AI systems by public authorities, as concerns for public values and fundamental rights of individuals are different. From a general point of view, AI can be used for cognitive purposes and auxiliary tasks or decision-making purposes⁸.

In current law enforcement activities, AI is used exclusively for support tasks. However, while it is generally acknowledged that the use of AI for decision-making raises serious concerns due to the replacement of the human being by a machine⁹, the use of AI systems to perform auxiliary tasks can also pose challenges that are far from easy to solve. An example is the well-known COMPAS case¹⁰, in which a proprietary predictive software was used by judges to determine the sentence of Eric L. Loomis, identified by the software as a high-risk recidivist. Loomis' defence asserted that the use of COMPAS in sentencing violated his right to a fair trial and appealed the decision¹¹. Although the use of COMPAS was not considered unconstitutional by the Supreme Court of Wisconsin, as the results of the software were not decisive for the decision¹², this case highlighted that even when predictive

algoritmi ed i pericoli del cambiamento in atto, in *Diritto dell'Economia*, 3, 2020, 599-634. A. DI MARTINO, *Tecnica e potere nell'amministrazione per algoritmi*, Editoriale Scientifica, 2023. See also D.-U. GALETTA, J.G. CORVALÁN, *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *Federalismi.it*, 3, 2019, 1-23.

⁶ John McCarthy described AI as “the science and engineering of making intelligent machines, especially intelligent computer programs”. J. MCCARTHY, *What is Artificial Intelligence?*, Stanford University, 2007. Available at: <https://www-formal.stanford.edu/jmc/whatisai.pdf>.

⁷ MARCHETTI, *Artificial Intelligence and Public Authorities: Does the European AI Act Protect Public Values?*, cited; MARCHETTI, *Amministrazione digitale*, cited, 78.

⁸ MARCHETTI, *Amministrazione digitale*, cited, 98.

⁹ For a discussion of the (in)compatibility of using AI for decision-making purposes in the context of justice see B. MARCHETTI, *Giustizia amministrativa e transizione digitale. Spunti per riflettere su un futuro non troppo lontano*, in M. RAMAJOLI (Ed.), *Una giustizia amministrativa digitale?*, Il Mulino, 2023, 59-90.

¹⁰ Supreme Court of Wisconsin, *State of Wisconsin v. Eric L. Loomis*, Case no. 2015AP157-CR, 2016.

¹¹ The violation of the right to a fair trial was contested under three main grounds: i) the right to be sentenced to a particular punishment on the basis of accurate information, (ii) the right to be sentenced to an individualised punishment; (iii) the improper use of gender data in the determination of the punishment. The defendant contested a violation of the *Due Process Clause* under the Fifth and Fourteenth Amendments of the U.S. Constitution.

¹² The sentence of the Supreme Court of Wisconsin concluded that software, such as COMPAS, can be used in sentencing, but risk scores (outcomes) derived from an automated system cannot be used as determining factors in deciding whether or not a convicted person should be imprisoned. The results of the software must be just one of the factors that judges must consider, along with many others, when making sentencing decisions.

tools are used for decision support, they may raise concerns, especially in relation to some public values, such as fairness¹³ and transparency¹⁴.

In Chapter 1, I pointed out that the protection of public order and security are assets of primary interest to the State. The protection of these interests therefore requires a restriction of the rights and freedoms of all those who these goods endanger, with important consequences on the protection of individuals, their constitutional rights and freedoms, and on the maintenance of public values, which serve as an indispensable guarantee for the individual.

Following a brief overview of the digitisation of police organisations, this Chapter examines the use of AI within law enforcement through the analysis of two paradigmatic applications. Specifically, the first part of the Chapter (Part I) focuses on predictive policing, scrutinising the epistemological assumptions underpinning predictive models, their operational deployment across selected European jurisdictions, and the resulting implications for fundamental rights. The second part of the Chapter (Part II) addresses AI-enabled video surveillance and face recognition technologies, analysing their technical capabilities, patterns of deployment, and their impact on privacy, data protection, freedom of assembly, and the principle of equality.

The Chapter demonstrates that, while these technologies are frequently justified on the basis of enhanced efficiency, effectiveness, and improved allocation of law enforcement resources, their use simultaneously entails structural and systemic risks to fundamental rights. These risks underscore the necessity for robust, context-specific regulatory safeguards capable of adequately constraining the deployment of AI systems in law enforcement settings.

¹³ J. ANGWIN, J. LARSON, S. MATTU, L. KIRCHNER, *Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks*, ProPublica, 23 May 2016. Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

¹⁴ The defence argued that it was impossible to scientifically refute the use of the data processed by COMPAS because intellectual property rights would have prevented the defence from accessing the information and mechanisms concerning the functioning of the software. However, the court held that Loomis had the possibility to challenge the final results of the risk calculation because the COMPAS manual explains how the scores are calculated. However, this part of the judgment raises serious doubts as to whether the defence is actually in a position to challenge the systems results. See Northpointe Inc., *Practitioners Guide to COMPAS Core*, 4 April 2019. Available at: <https://archive.epic.org/algorithmic-transparency/crim-justice/EPIC-16-06-23-WI-FOIA-201600805-COMPASPractitionerGuide.pdf>. For an interesting analysis of COMPAS, see C. RUDIN, C. WANG, B. COKER, *The Age of Secrecy and Unfairness in Recidivism Prediction*, in *Harvard Data Science Review*, 2(1), 2020, 1-53.

2. The digitisation of police organisations

Simon Egbert and Matthias Leese introduced the concept of “platformization of police work”¹⁵ to describe the transformation process that policing is undergoing through data and digital technologies. Technological tools have always played a key role in all areas of police work, such as mobility, surveillance, detection and analysis¹⁶. Examples of technological tools used by police authorities are crime mapping software, computerised record management systems, body-worn or car cameras, number plate readers and data mining software. However, there is a considerable variation in techniques used, in the sophistication of tools and methods and in the frequency of their uses among different police organisations. This variation can be found both between different countries and between different departments within the same country, due to historical and political reasons and national and local specificities¹⁷.

Before turning to the current frontier of AI in policing, it seems necessary to delve into the complex longstanding relationship between policing and technology. The discourse on technology and policing inevitably requires looking, first and foremost, at the United States, which have always led the way on the subject. Europe normally follows, with a few years delay, the progress made by the United States in relation to crime control.

From a general perspective, the digitisation process of the police can be traced back to the mid-1960s in the U.S. when the demand for better policing led to the identification of technology as an indispensable tool for carrying out police work¹⁸. The adoption of computers and the digitisation of

¹⁵ S. EGBERT, M. LEESE, *Criminal Futures. Predictive Policing and Everyday Police Work*, Routledge, 2021, 11. S. EGBERT, *Predictive Policing and the Platformization of Police Work*, in *Surveillance & Society*, 17 (1/2), 2019, 83-88. The author writes that “*The platformization of police work is understood as an organizational process in which manifold data sets and databanks—especially from police-external sources—are cross-linked, creating information retrieval and production networks designed to improve police work on numerous levels by facilitating knowledge creation (e.g., patrol allocation, police management, crime investigation, etc.). Ibidem*, 84. Similarly, Sara Brayne talks about “datafication of policing” or «policing by numbers». S. BRAYNE, *Predict and Surveil. Data, Discretion, and the Future of Policing*, Oxford University Press, 2020. Kelly Gates talks about “platform policing”. K. GATES, *Policing as Digital Platform*, in *Surveillance & Society*, 17 (1/2), 2019, 63-68.

¹⁶ Among others, see S.D. MASTROFSKI, J.T. WILLIS, *Police Organization Continuity and Change: Into the Twenty-First Century*, in *Crime and Justice*, 39(1), 2010, 79-80; P.K. MANNING, *Policing Contingencies*, The University of Chicago Press Ltd, 2003.

¹⁷ The United States are far ahead on the topic. See E. MAGUIRE, *Organizational Structure in American Police Agencies: Context, Complexity, and Control*, SUNY Press, 2003. Some interesting reports do exist. See RTI International Police Executive Research Forum, *Research on the Impact of Technology on Policing Strategy in the 21st Century*, May 2016. As far as Europe is concerned, see F. JANSEN, *Data Driven Policing in the Context of Europe*, Working Paper, *Datajustice*, 2018.

¹⁸ See President’s Commission on Law Enforcement and Administration of Justice, *The Challenge of Crime in a Free Society*, Report, 1967. Available at: <https://www.ojp.gov/sites/g/files/xyckuh241/files/archives/ncjrs/42.pdf>. For a first discussion on the use of technology by the U.S. law enforcement authorities see K.W. COLTON, *The Impact and Use of Computer Technology by the Police*, in *Communications of the ACM*, 22(1), 1979, 10-20.

CHAPTER 2

police records are considered the first step in the digitisation process¹⁹. However, while the discourse on technology in policing has been extensive²⁰, far less attention has been devoted to its practical implementation. The introduction of technology into police work was indeed a slow process, shaped both by the behavioural and power dynamics at stake and by persistent questions surrounding the capacity of technological tools to meet their stated objectives, their ultimate impact, and the balance between their benefits and costs²¹.

It is generally acknowledged that policing has always been a complex setting for innovation, since it is highly influenced by social, economic and political factors and deeply resistant to rapid change²². Policing is generally characterised by traditional long-standing strategies, tactics and practices and innovation is not always perceived as something positive, with different reactions within the same organisation²³. Consequently, although technology has often been credited with increasing efficiency and effectiveness and improve crime-control, research has also shown that technology can have unintended and negative consequences on these same goals²⁴. As scholars pointed out, technological systems, when they enter police organisations, cannot be considered as mere technical tools to improve effectiveness and efficiency, but become socio-technical systems that involve a multiplicity of relationships and interactions with the environment in which they are used. The introduction of technology in policing is more of a “social process”²⁵ that requires a broader analysis of how the society and the organisation itself responds to it²⁶, and it is precisely this response that could lead to an improvement in the effectiveness and efficiency of the organisation’s performance²⁷.

Despite an initial inertia on the part of police forces, technology has started to grow, albeit moderately, within police departments. The 1990s represented a flourishing period for the digitisation of police

¹⁹ M. AFZAL, P. PANAGIOTOPOULOS, *Data in Policing: An Integrative Review*, in *International Journal of Public Administration*, 48(7), 2025, 411-430. A. NORTHROP, K.L. KRAEMER, J.L. KING, *Police use of computers*, in *Journal of Criminal Justice*, 23(3), 1995, 259–275.

²⁰ One can detect a dichotomy in the discourse: at one extreme, there are those who view technology in policing with caution, emphasising the possible danger of technological progress in such a sensitive area. At the other extreme, there are those who support the benefits of technology in policing and believe these outweigh the costs of its implementation.

²¹ COLTON, *The Impact and Use of Computer Technology by the Police*, cited, 10.

²² See D. ALLEN, S. KARANASIOS, *Critical factors and patterns in the innovation process*, in *Policing (Oxford): a journal of policy and practice*, 2011, 5(1), 87-97.

²³ See BRAYNE, *Predict and Surveil*, cited, 74. See also P.K. MANNING, *Technological Dramas and the Police: Statement and Counterstatement in Organizational Analysis*, in *Criminology*, 30(3), 1992, 328.

²⁴ See C. LUM, C.S. KOPER, J. WILLIS, *Understanding the Limits of Technology’s Impact on Police Effectiveness*, in *Police Quarterly*, 2017, 20(2), 135–163. P.K. MANNING, *Technological Dramas and the Police*, cited.

²⁵ F. JANSEN, *Data-driven policing: Negotiating the legitimacy of the police*, PhD Thesis, Department of Journalism, Media and Culture Cardiff University, January 2022, 10. Available at: <https://orca.cardiff.ac.uk/id/eprint/149904/1/2022jansenjphd.pdf>.

²⁶ Currently, there are not many studies on how working platforms are perceived by the authorities using it and by potential recipients. Among those that do exist, an interesting analysis concerns Estonia and Sweden, see A. MASSO, T. KASAPOGLU, A. KAUN, V. GALIS, *Citizens’ perspectives on platformisation of police work: a scenario and story-based exploration in Estonia and Sweden*, in *Information, Communication & Society*, 27 (13), 2024, 2400–2418.

²⁷ EGBERT, LEESE, *Criminal futures*, cited, 53.

CHAPTER 2

work²⁸. The new managerialism, that introduced private sector theories, techniques, tools and practices into the governance structure of the public sector, gave a strong impetus to the adoption of technology in policing²⁹. Janet B.L. Chan defined “three imperatives”³⁰ that supported this process³¹: i) improve effectiveness and efficiency, ii) satisfy the demands of external agencies for information and iii) meet the requirements of new forms of police management and accountability. The adoption of technological tools has thus allegedly led to a “professionalization of policing”³² or to a rationalization of policing³³, whose basic assumption was that the use of digital technologies brings about a shift towards a more scientific approach to policing³⁴, which is credited with leading to better results in terms of crime reduction and making policing more objective, transparent and accountable³⁵.

The attacks of 11 September 2001 intensified the use of digital technologies in policing, led to an increase in data available to police departments and to a broader acceptance of police’s use of technologies by the community³⁶. In fact, federal agencies have started funding local law enforcement authorities to collect, analyse and share huge volumes of data to inform decision-making³⁷. In addition, advances in computer power and data analysis techniques for the inspection, cleaning, transforming and modelling of data have further enhanced the knowledge that can be gained from large amounts of

²⁸ S. NUNN, *Police Information Technology: Assessing the Effects of Computerization on Urban Police Functions*, in *Public Administration Review*, 61(2), 2001, 221–234.

²⁹ «Managerialism: A term associated with the shift in government policy towards ‘new public management’ characterised by, inter alia, elements of privatisation; marketization; the increased use of performance indicators; a growing emphasis on outputs and outcomes; partnership working; and the redesignation of clients as ‘customers’». T. NEWBURN, *Handbook of policing*, Willan Publishing, 2003, 728. The new managerialism was highly criticised because it focused mainly on performance, outputs and outcomes and for giving more attention to quantity over quality. For a critical view of the new managerialism in the police organisation see A. BECKLEY, *Is Managerialism Alive and Well in the Police Service?*, in J.F. ALBRECHT, G. DEN HEYER (Eds.), *Enhancing Police Service Delivery*, 2021, 341–358. J.B.L. CHAN, *Governing police practice: limits of the new accountability*, in *British Journal of Sociology*, 50(2), 1999, 251–270.

³⁰ J.B.L. CHAN, *The Technological Game: How Information Technology is Transforming Police Practice*, in *Criminal Justice*, 1(2), 2001, 140. Similarly, J.B.L. CHAN, D. BRERETON, M. LEGOSZ, S. DORAN, *E-policing: The Impact of Information Technology on Police Practices*, 2001, 3–4. Available at: <https://www.ojp.gov/ncjrs/virtual-library/abstracts/e-policing-impact-information-technology-police-practices>.

³¹ Similarly, R.V. ERICSON, K.D. HAGGERTY, *Policing the risk society*, University of Toronto Press, 1997. P.K. MANNING, *Information Technologies and the Police*, in *Modern Policing*, 15, 1992, 349–398. EGBERT, LEESE, *Criminal futures*, cited, 45.

³² BRAYNE, *Predict and Surveil*, cited, 18.

³³ This «rationalisation» emerges from the work P.K. MANNING, *Information Technologies and the Police*, cited.

³⁴ Sarah Brayne gives credit to August Vollmer, Chief of the Berkeley Police Department, for promoting the application of scientific principles in policing. The author describes Vollmer as the «father of modern law enforcement». BRAYNE, *Predict and Surveil*, cited, 19. The concept of «scientification» of police work can be found in R.V. ERICSON, C.D. SHEARING, *The Scientification of Police Work*, in G. BÖHME, N. STEHR (Eds.) *The Knowledge Society: The Growing Impact of Scientific Knowledge on Social Relations*, Reidel, 1986, 129–159.

³⁵ EGBERT, LEESE, *Criminal futures*, cited, 48.

³⁶ BRAYNE, *Predict and Surveil*, cited, 21. S. BRAYNE, *Big Data Surveillance: The Case of Policing*, in *American Sociological Review*, 82(5), 2017, 977–1008.

³⁷ Local law enforcement authorities were seen as actors on the front lines of the domestic war against terrorism. See M.C. WAXMAN, *Police and National Security: American Local Law Enforcement and Counter-Terrorism after 9/11*, in *Journal of National Security Law and Policy*, 3, 2009, 377–407.

data held by the police³⁸. The last decades have thus seen the rise of integration of data-driven approaches in policing, giving rise to new data-driven policing paradigms³⁹. The police's turn to data-driven approaches is currently evident also in European police organisations⁴⁰.

It is well known that the police collect and store data every day, through everyday patrolling, stopping and questioning people, calls for services, police reports, information collected during investigations and criminal justice databases. In addition, the “datafication”⁴¹ of our society also increases the amount of data available which can be retrieved and analysed by the police⁴². In their enlightening work, Ian H. Witten, Eibe Frank and Mark A. Hall underline the growing gap between the generation of data and our understanding of it: the more data is produced, the less we understand⁴³. However, as they point out, hidden in all this data rests potentially useful information: if analysed intelligently, the data can be considered an extremely valuable resource. Currently, the development of increasingly complex algorithms for data analysis allows data to be exploited in new ways, discovering patterns, trends, links that can be later used to support decision-making.

Analysing a large amount of data nowadays has become a crucial aspect of efficient and effective decision-making in many sectors⁴⁴. Law enforcement is no exception. Indeed, data analysis has always been central to law enforcement, as authorities operate with limited budgets and personnel in time-sensitive settings that demand real-time information for sound decision-making⁴⁵. The stakes in this area are particularly high: public safety is a fundamental and constitutional interest, and flawed decisions by law enforcement can directly undermine citizens' rights and freedoms⁴⁶.

³⁸ AFZAL, PANAGIOTOPOULOS, *Data in Policing*, cited, 6.

³⁹ *Ibidem*. BRAYNE, *Big Data Surveillance: The Case of Policing*, cited. An interesting analysis of Big Data and policing was conducted by A. BABUTA, *Big Data and Policing An Assessment of Law Enforcement Requirements, Expectations and Priorities*, RUSI Occasional Paper, September 2017. Available at: [201709_rusi_big_data_and_policing_babuta_web.pdf](https://rusi.org/publications/201709_rusi_big_data_and_policing_babuta_web.pdf).

⁴⁰ See JANSEN, *Data Driven Policing in the Context of Europe*, cited.

⁴¹ The term was introduced by Kenneth Cukier and Viktor Mayer-Schönberger in 2013. The authors describe «datafication» as the reduction of any phenomenon into a quantified format that can be recorded and retrieved, tabulated and analysed. K. CUKIER, V. MAYER-SCHÖNBERGER, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, 2013.

⁴² J. VAN DIJCK, *Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology*, in *Surveillance & Society*, 12(2), 2014, 197-208.

⁴³ I.H. WITTEN, E. FRANK, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2011, 4.

⁴⁴ Currently, most uses of AI by public administrations are not for decision-making but to support decision-making. In fact, public administrations use AI systems to extract knowledge from data and use this knowledge to inform subsequent actions, such as controls on citizens or private companies.

⁴⁵ The need for swift and effective decision-making is evident both in the area of domestic security but is even more evident in relation to external security, with respect to terrorist attacks, where swift action is essential to detect and avert the danger. The importance of information as the basis for police's actions emerges from P.K. MANNING, *Information Technologies and the Police*, cited.

⁴⁶ In some cases, it has even been argued that there is an obligation on the police to innovate and that this obligation is directly related to the police's public protection duty and positive responsibilities under Articles 2 and 3 of the ECHR. For a discussion on the topic see A. BABUTA, M. OSWALD, *Data Analytics and Algorithms in Policing in England and Wales: Towards A New Policy Framework*, RUSI Occasional Paper, February 2020, 7 reporting what was said in a Roundtable event organised by RUSI and CDEI, London, 25 July 2019.

Data-driven technologies now stand at the forefront of modern law enforcement, with AI-powered analysis representing the field's cutting edge⁴⁷. This turn to data and AI-driven approaches further underlines the paradigm shift discussed in Chapter 1, which concerns the shift from a reactive to a preventive approach in the fight against crime⁴⁸. The shift from a repressive to a preventive approach in policing is perfectly compatible with the “risky society”⁴⁹, described by Ulrich Beck, who claimed that the production and management of risk is the defining feature of contemporary societies. In a risk society, certainty gives way to pervasive uncertainty, uncertainty breeds insecurity, which in turn fuels the demand to preempt harmful events before they occur⁵⁰. Therefore, a demand for risk prevention arises and requires institutions that derive their authority from controlling such hazards to act before the risk itself occurs⁵¹. Translating this logic into policing, crime breeds insecurity and increases the demand on the police not only to activate once there is a suspicion that a crime will be committed or once the crime has been committed, but also to eliminate the very risk of crime occurring.

Data-driven policing is precisely directed towards this goal⁵². The idea on which data-driven policing is based is that of *prediction*, as I have described it in Chapter 1, and predictive analytics, which is the foundation of the preventive approach⁵³. Indeed, prevention involves taking proactive steps to avoid or reduce the likelihood of negative events based on the insights from predictive analytics. In relation to predictive policing, for example, police departments can use predictive software to analyse data, predict where and when a certain type of crime is more likely to occur, and, once the police have identified the crime hotspots, take preventive actions to reduce the risk of crime occurring. Some authors describe this attitude as “a shift from perception of crime as harm to a perception of crime as risk”⁵⁴ leading to a move from harm prevention to risk prevention⁵⁵. Doubts about the compatibility

⁴⁷ See AFZAL, PANAGIOTOPOULOS, *Data in Policing*, cited for a panorama of possible uses of data-driven technologies in policing. See also JANSEN, *Data Driven Policing in the Context of Europe*, cited. Europol, *AI and policing. The benefits and challenges of artificial intelligence for law enforcement*, Report from the Europol Innovation Lab, 2024.

⁴⁸ R. URSI, *La sicurezza nelle città: dallo spazio difendibile a tecnopolis*, in *Munus*, 2, 2024, 1-31.

⁴⁹ U. BECK, *Risk Society: Towards a New Modernity*, Sage Publications, 1992.

⁵⁰ R. URSI, *La sicurezza nelle città*, cited, 3. Risk «Is a concept that presupposes decisions that attempt to make the unpredictable consequences of civilization's decisions predictable and controllable». Own translation from M. CARASSAI, *Paura e società del rischio. Un'intervista a Ulrich Beck*, in *Lo Sguardo*, 21, 2016, 210 which represents the Italian translation of the text *An Interview with Ulrich Beck on Fear and Risk Society*, in *The Hedgehog Review*, Fall 01, 96-107.

⁵¹ R. URSI, *La sicurezza nelle città*, cited, 3.

⁵² For a critical consideration on actuarial methods in the area of criminal justice and law enforcement see B.E. HARCOURT, *Against Prediction. Profiling, Policing, and Punishing in an Actuarial Age*, University of Chicago Press, 2008.

⁵³ For a discussion on the reliance on predictions in policing see C. ARADAU, T. BLANKE, *Politics of prediction: Security and the time/space of governmentality in the age of big data*, in *European Journal of Social Theory*, 20(3), 2016, 373-391.

⁵⁴ L. STRIKWERDA, *Predictive policing: The risks associated with risk assessment*, in *The Police Journal: Theory, Practice and Principles*, 94(3), 2021, 426. The idea of crime as risk instead of harm is based on L. ZEDNER, *Pre-crime and post-criminology*, in *Theoretical Criminology*, 11(2), 2007, 261-281.

⁵⁵ An interesting book which deals with society and the prevention of risk is T. PITCH, *La società della prevenzione*, Carocci, 2007.

of this approach with the constitutional paradigm may arise⁵⁶. In fact, to prevent the risk of crime occurring, it is necessary to anticipate the moment of activation of police activity to a stage when there is still a lack of an unambiguous manifestation that legitimises a suspicion. The result is an inevitable weakening of constitutional and legislative guarantees for the individual, which generally provide protection only from the moment of the emergence of a suspicion or even after the crime is committed. Some scholars have gone further, speculating that crime risk management through the use of data-driven technologies has led to the so-called “control culture”⁵⁷ described by David Garland, as a global phenomenon that pervades contemporary societies, shifting towards a digital surveillance logic and leading to the rise of a “surveillance state”⁵⁸ where individuals are constantly under control⁵⁹. Beyond these surveillance studies⁶⁰, it is clear that the entry of data-driven technologies into policing has several consequences⁶¹. Advanced digital technologies, and AI in particular, are pervasive and transformative in nature⁶², profoundly transforming the way the police work and how they are perceived by the community⁶³. This may open up concerns in terms of privacy and data protection, as well as chilling effects on the exercise of fundamental rights, such as freedom of assembly and freedom of expression⁶⁴.

Among the many data-driven technologies, for the purpose of this Thesis I will focus on predictive policing and AI-enhanced video analysis, with a focus on face recognition technologies. I chose these case studies because both were at the centre of the debate between the European institutions during the adoption of the AI Act, leading to opposing views on these applications up to the final version. Moreover, they currently represent two of the most critical applications in the context of law enforcement regarding domestic public safety.

⁵⁶ In these terms see URSI, *La sicurezza nelle città*, cited, 22. See also the book review O. DE LEONARDIS, *T. Pitch, La società della prevenzione*, 2007, in *Rassegna Italiana di Sociologia*, 1, 2008, 131-133. In the U.S. context similar concerns arose, questioning the compatibility of predictive policing with the Fourth Amendment reasonable suspicion. See A.G. FERGUSON, *Illuminating Black Data Policing*, in *Ohio State Journal Of Criminal Law*, 15, 2018, 520.

⁵⁷ D. GARLAND, *The Culture of Control. Crime and Social Order in Contemporary Society*, University of Chicago Press, 2001. In the book, Garland describes in an extremely interesting way the reconfiguration of the response to crime and the sense of criminal justice in modern societies.

⁵⁸ J. CHIN, L. LIN, *Stato di sorveglianza. La via cinese verso una nuova era del controllo sociale*, Bollati Boringhieri, 2022.

⁵⁹ One of the most recognised books on the subject is the masterpiece S. ZUBOFF, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, PublicAffairs, 2019.

⁶⁰ P. FUSSEY, A. SANDHU, *Surveillance Arbitration in the Era of Digital Policing*, in *Theoretical Criminology*, 26(1), 2021, 3–22.

⁶¹ For general discussion on the power of algorithms in relation to social processes see D. BEER, *Introduction: The social power of algorithms*, in D. BEER (Ed.), *The social power of algorithms*, Routledge, 2018.

⁶² For an interesting discussion on the changes brought by technologies see N. POSTMAN, *Technopoly. The Surrender of Culture to Technology*, Vintage Books, 1992.

⁶³ See CHAN ET AL., *E-policing: The Impact of Information Technology on Police Practices*, cited. J.B.L. CHAN, *The technological game: how information technology is transforming police practice*, in *Criminal Justice*, 1(2), 2001, 139-159.

⁶⁴ An interesting study on AI in law enforcement and the impact on fundamental rights was requested by the LIBE committee. G. GONZÁLEZ FUSTER, *Artificial Intelligence and Law Enforcement. Impact on Fundamental Rights*, July 2020. Available at: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU\(2020\)656295_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf).

Part I: Predictive policing

1. The current frontier of data analytics: predictive policing

The objective of this section is to address in depth the topic of predictive policing, which is currently one of the most discussed uses of data analytics in law enforcement. I will proceed through empirical research, starting from a brief overview of predictive policing and focusing on current practices in Europe, where this field is still little studied⁶⁵. The reference to the United States is also necessary because advanced algorithmic tools are now being used in several States to inform or to assist human decision-making, and insights can be drawn from the vast literature on the subject⁶⁶. This paragraph does not intend to provide an extensive overview of all the tools that are being developed and used across Europe, but rather to approach them as case studies that offer interesting insights for answering the research questions.

1.1. The experience of predictive policing in the United States

CompStat (short for Computer Statistics or Comparative Statistics)⁶⁷ has traditionally been considered the predecessor of predictive policing. Its origins can be traced back to the early 1990s when the Police Commissioner Bill Bratton and his deputy Commissioner for crime control strategies, Jack Maple, reoriented the New York Police Department (NYPD) to focus on crime prevention instead of crime suppression and developed and implemented the new data-driven system that was eventually called CompStat⁶⁸. CompStat was described as a performance management system⁶⁹ and a methodology⁷⁰ for dealing with crime based on the processing and analysis of crime statistics and regular meetings between department managers and officers to discuss crime-related problems and strategies for solving

⁶⁵ Interesting research on the topic was recently published. See JANSEN, *Data-driven policing*, cited.

⁶⁶ The literature on the topic is huge. In addition to those already mentioned, see B. BENBOUZID, *To Predict and to Manage: Predictive Policing in the United States*, in *Big Data & Society*, 6, 2019, 1–13. W.J. BRATTON, S.W. MALINOWSKI, *Police Performance Management in Practice: Taking COMPSTAT to the Next Level*, in *Policing: A Journal of Policy and Practice*, 2(3), 2008, 259–265. W. BRATTON, P. KNOBLER, *Turnaround. How America's Top Cop Reversed the Crime Epidemic*, Random House, 1998.

⁶⁷ J. MAPLE, C. MITCHELL, *The Crime Fighter: How You Can Make Your Community Crime-Free*, Broadway Books, 2000.

⁶⁸ W.J. BRATTON, S.W. MALINOWSKI, *Police Performance Management in Practice: Taking COMPSTAT to the Next Level*, in *Policing*, 2008, 2(3), 259–265.

⁶⁹ *Police Leaders at PERF/BJA Meeting Discuss CompStat: Best Practices and Future Outlook*, in *Newsletter Of The Police Executive Research Forum*, March/April, 25(2), 2011. Available at: https://www.policeforum.org/assets/docs/Subject_to_Debate/Debate2011/debate_2011_marapr.pdf.

⁷⁰ D. WEISBURD, S. MASTROFSKI, J.J. WILLIS, R. GREENSPAN, *Changing everything so that everything can remain the same: Compstat and American policing*, in D. WEISBURD, A.A. BRAGA (Ed.), *Police Innovation: Contrasting Perspectives*, Cambridge University Press, 2001, 284–301.

them⁷¹. To this end, police units were required to collect and analyse data on the unit's past performance, follow up on previous decisions and commitments to improve performance, set new performance goals and examine the effectiveness of performance strategies⁷². After its implementation, CompStat brought multiple benefits to the NYPD: it contributed to information sharing by facilitating the flow of information between divisions, decentralised problem solving by providing more authority to commanders, increased accountability and brought about a change in the organisational structure of police agencies⁷³. General opinion was therefore predominantly optimistic⁷⁴ and after its success at the NYPD the methodology was adopted by other police⁷⁵ and non-police⁷⁶ agencies, each with its own specificities⁷⁷.

The use of past data to inform future strategies formed the basis of the subsequent predictive policing approaches that arose from the convergence in the 2000s of criminological and policing theories⁷⁸. In relation to criminological theories, the most relevant criminal behaviour theories are the theory of the rational choice, the routine activity theory and the crime pattern theory⁷⁹. In relation to policing theories, on the other hand, the reference goes mainly to problem-oriented policing (POP)⁸⁰ and

⁷¹ Bureau of Justice Assistance, Police Executive Research Forum, *Compstat: Its Origins, Evolution, and Future In Law Enforcement Agencies*, 2013, 2. Available at: <https://bja.ojp.gov/sites/g/files/xyckuh186/files/Publications/PERF-Compstat.pdf>.

⁷² J.J. WILLIS, S.D. MASTROFSKI, D. WEISBURD, *Compstat in Practice: An In-Depth Analysis of Three Cities*, Police Foundation, 2003. This report is the third stage of a project, entitled "*Compstat and Organizational Change: A National Assessment*", funded by the National Institute of Justice and conducted by the Police Foundation. CompStat was based on four main principles developed by NYPD Deputy Commissioner Jack Maple: i) accurate and timely intelligence, ii) effective tactics, iii) rapid deployment and iv) relentless follow-up.

⁷³ Mainly introducing flexibility and creativity. See BJA, *Compstat*, cited, 5-6.

⁷⁴ See C. KRAUSS, *Crime Lab: Mystery of New York, the Suddenly Safer City*. Article published on the *New York Times*, 27 July 1995. Available at: <https://www.nytimes.com/1995/07/23/weekinreview/crime-lab-mystery-of-new-york-the-suddenly-safer-city.html>.

⁷⁵ For example, CompStat was implemented in the Los Angeles Police Department when Bill Bratton became Police Chief in this Department and Garry McCarthy brought CompStat in the Chicago Police Department.

⁷⁶ An example is PerformanceStat developed by Professor Robert Behn. R.D. BEHN, *The PerformanceStat Potential. A Leadership Strategy for Producing Results*, Brookings Institution Press, 2014.

⁷⁷ Although CompStat was generally evaluated positively, some resistance can be identified. See BJA, *Compstat*, cited, 15-16.

⁷⁸ G.M. CAMPEDELLI, *Machine Learning for Criminology and Crime Research: At the Crossroads*, Routledge, 2022, 52.

⁷⁹ R.V. CLARKE, M. FELSON (Eds.), *Routine Activity and Rational Choice (Advances in Criminological Theory)*, Routledge, 2018. L.E. COHEN; M. FELSON, *Social Change and Crime Rate Trends: A Routine Activity Approach*, in *American Sociological Review*, 44(4), 1979, 588–608. P.J. BRANTINGHAM, P.L. BRANTINGHAM, M.A. ANDRESEN, *The geometry of crime and crime pattern theory*, in R. WORTLEY, M. TOWNSLEY (Eds.), *Environmental criminology and crime analysis* (2nd Ed.), New York, Routledge, 2017, 98-115. D.B. CORNISCH, R.V.G. CLARKE (Eds.), *The Reasoning Criminal: Rational Choice Perspectives on Offending*, Routledge, 2014. M. FELSON, L.E. COHEN, *Social Change and Crime rate Trends: A Routine Activities Approach*, in *American Sociological Review*, 44, 1979, 588-608.

⁸⁰ The concept of problem-oriented policing was articulated and later elaborated by Herman Goldstein, a professor at the University of Wisconsin Law School in Madison. This concept is based on the idea that the police, to carry out its activities effectively, must identify crime-related problems in the community, explore these problems, namely, to inquire them, acquire the necessary information and analyse it, and finally identify appropriate responses to each problem, exploring different alternatives. The success of problem-oriented policing strategies is due to the recognised limitations of typical police operations (e.g., random patrols, response to a committed crime, criminal investigations) which are not always effective in solving the wide range of community problems that the police is expected to handle, many of which are not strictly criminal in nature. Goldstein articulated and elaborated the concept of Problem-Oriented Policing in two seminal

evidence-based policing (EBP)⁸¹. Slightly different, because of the type of crime it targets, is intelligence-led-policing (ILP)⁸². The core idea on which those theories are based is crucial for contemporary law enforcement: operational and tactical decisions in law enforcement must be based on the systematic collection, organisation and analysis of data. To simplify, all police records are *data*, which are then analysed to become *information*. The information is then interpreted and understood by the police officer, becoming *knowledge*. Once more data and information are collected and once the data, information and knowledge are evaluated, analysed and presented in a decision-making format, they become *intelligence*, which is the basis for setting strategic and operational priorities and objectives in crime prevention and suppression⁸³.

Based on these police theories, predictive policing has begun to emerge. In late 2000s the National Institute of Justice, in partnership with the Bureau of Justice Assistance and the Los Angeles Police Department (LAPD) held a Predictive Policing Symposium which represented one of the first occasions where researchers and practitioners gathered to discuss about this new concept⁸⁴. In 2010, a second Predictive Policing Symposium was held in Rhode Island, further extending the discussion of

publications. H. GOLDSTEIN, *Improving Policing: A Problem-Oriented Approach*, in *Crime and Delinquency*, 25(2), 1979, 236-258. H. GOLDSTEIN, *Problem-Oriented Policing*, McGraw-Hill, 1990. In his seminal paper Goldstein stresses that the police job can be better described as dealing with problem instead of simply enforcing the law, being the latter only a means to an end. GOLDSTEIN, *Improving Policing*, cited, 242. For a discussion on the positive aspects of problem-oriented policing see Police Executive Research Forum, *Excellence in Problem-Oriented Policing: The 2001 Herman Goldstein Award Winners*, 2011. Available at: <https://popcenter.asu.edu/sites/default/files/library/awards/goldstein/2001/2001.pdf>. The SARA (Scanning-Analysis-Response-Assessment) problem-solving model, which represents the most used process for addressing recurring problems, builds on the problem-oriented policing project in Newport News Police Department of mid-1980's. See J. ECK, W. SPELMAN, *Problem solving: Problem-oriented policing in Newport News*, Police Executive Research Forum, 1987. Available at: <https://www.ojp.gov/pdffiles1/Digitization/111964NCJRS.pdf>.

⁸¹ The paradigm of evidence-based policing, on the other hand, was introduced by Lawrence W. Sherman and focuses on the fundamental idea that «*police practices should be based on scientific evidence about what works best*». It aims at stressing a link between research and practice, using best available research on the outcomes of police work to guide practice and evaluate police agencies, units and officers. The main idea was thus to bring the scientific approach into the policing sector. L.W. SHERMAN, *Evidence-Based Policing*, in *Ideas in American Policing*, 1998. In presenting this new paradigm, Sherman tried to answer different questions: What is evidence-based policing? What is new about it? How does it apply to a specific example of police practice? How can it be institutionalized?. There is a strong connection between problem-oriented policing and evidence-based policing, as the evaluation of problem-solving responses is a fundamental part of problem-oriented policing and the core idea of evidence-based policing, but the latter goes a step further by introducing the concept of scientific evidence for the selection of strategies and the monitoring of their implementation. One of the works that contributed significantly to the establishment of this new way of analysing crime was *Become a Problem-Solving Crime Analyst* by Ronald Clarke and John E. Eck, which proposed a clear method to analyse crime based on fifty-five steps. R. CLARKE, J.E. ECK, *Become a Problem-Solving Crime Analyst*, Routledge, 2003. In their book the authors proposed the SARA (Scanning-Analysis-Response-Assessment) model.

⁸² While the previous models generally focus on local areas and threats, intelligence-led-policing is a tool for combat transnational organised crime, countering terrorism and violent extremism and radicalization that lead to terrorism (so called «VERLT »). J.H. RATCLIFFE, *Intelligence-Led Policing*, Routledge, 2016.

⁸³ OSCE, *Guidebook Intelligence-Led Policing*, 2017. Available at: <https://www.osce.org/files/f/documents/d/3/327476.pdf>.

⁸⁴ National Institute of Justice, *Predictive Policing Symposiums*. Available at: <https://www.ojp.gov/pdffiles1/nij/242222and248891.pdf>.

the previous Symposium. From this moment on, the term predictive policing has entered the law enforcement circle.

The following years were a successful period for predictive policing programmes, which spread rapidly throughout the United States and dozens of departments began implementing them. Indeed, predictive policing has been presented as the promising future of law enforcement⁸⁵: it has the appeal of being based on empirical data, introducing the scientific approach into the police work; it has the appeal of being neutral because it relies on algorithms and it presents itself as the answer to the age-old problem of crime, namely anticipating and preventing crime before it has been committed⁸⁶.

The Santa Cruz Police Department was one of the first law enforcement agencies to implement a predictive policing program in 2011 named PredPol (short for Predictive Policing)⁸⁷. PredPol began as a collaborative project between the Los Angeles Police Department and the University of California, Los Angeles with the support of researchers from Santa Clara University⁸⁸. Later, it became a commercial software sold by PredPol Incorporation (now Geolitica), funded by those who initially participated in the project. As described in the *Best Practices and Training Guide*⁸⁹, one of the rare publicly available manual of a predictive policing software, PredPol is given crime data (what, when and where data) and this information is run through an algorithm that provides a prediction on hotspot locations⁹⁰ which can be used to inform the subsequent police operations and to allocate resources more effectively. Subsequently, the LAPD started using the programme and many other departments showed interest.

⁸⁵ Their aim was to «discuss the emerging framework and its impact on the future of law enforcement» and explore «the policy implications, privacy issues and technical elements of predictive policing». See B. PEARSALL, *Predictive Policing: The Future of Law Enforcement?*, in *National Institute of Justice Journal*, 266, 2010, 16-19.

⁸⁶ Multiple articles were published on different journals stressing the innovative strategies to combat crime. See E. HUET, *Server and Protect: Predictive Policing Firm PredPol Promises to Map Crime Before It Happens*. Article published in *Forbes*, 11 February 2015. Available at: <https://www.forbes.com/sites/ellenhuet/2015/02/11/predpol-predictive-policing/>. J. JOUVENAL, *Police are Using Software to Predict Crime. Is it a 'Holy Grail' or Biased Against Minorities?*. Article published in *The Washington Post*, 17 November 2016. Available at: https://www.washingtonpost.com/local/public-safety/police-are-using-software-to-predict-crime-is-it-a-holy-grail-or-biased-against-minorities/2016/11/17/525a6649-0472-440a-aae1-b283aa8e5de8_story.html. E. GOODE, *Sending the Police Before There's a Crime*. Article published in *The New York Times*, 15 August 2011. Available at: <https://www.nytimes.com/2011/08/16/us/16police.html>.

⁸⁷ Interestingly, Santa Cruz became the first city in the United States to put a moratorium on the use of predictive policing and eventually ban it. See K. STURGILL, *Santa Cruz becomes the first U.S. city to ban predictive policing*. Article published in the *Los Angeles Times*, 26 June 2020. Available at: <https://www.latimes.com/california/story/2020-06-26/santa-cruz-becomes-first-u-s-city-to-ban-predictive-policing>. Similarly, the Los Angeles Department ended some programs. J. BHUIYAN, *LAPD ended predictive policing programs amid public outcry. A new effort shares many of their flaws*. Article published in *The Guardian*, 8 November 2021. Available at: <https://www.theguardian.com/us-news/2021/nov/07/lapd-predictive-policing-surveillance-reform>.

⁸⁸ George Mohler was one of the researchers involved in the project.

⁸⁹ Predpol Inc., *Best Practices and Training Guide*. Revised 2014. Available at: <https://www.muckrock.com/foi/elgin-7770/foia-elgin-police-dept-predpol-documents-51858/#file-190432>.

⁹⁰ See Chapter 2, Section 1.2.1.

Outside the United States, PredPol has also found interested police agencies. The Kent Police, for example, sent a team to Los Angeles in 2012 to meet with LAPD commanders and conducted research on PredPol in order to implement the programme in the United Kingdom⁹¹. In 2014, the Kent Police provided a review of its evaluation of the programme after sixteen months of experimentation with overall positive conclusions⁹². However, in 2018 it seems that the Kent Police halted its deployment⁹³. Based on the extensive literature on the topic, it is possible to identify different predictive policing's phases of evolution⁹⁴. The first phase of predictive policing was directed towards targeting places of property crimes (e.g., burglaries and car thefts)⁹⁵. The second phase of predictive policing was mainly focused on targeting places of violent crimes (such as gun-related crimes or gang violence)⁹⁶. The third, and currently the last, phase of predictive policing is mainly concerned with targeting persons or groups involved in criminal activities, both as perpetrators and as victims.

1.2. Defining predictive policing

Before entering further details, it is crucial to define the concept of predictive policing. Indeed, it has emerged as a collective term that encompasses different ways in which the police analyse data through predictive analysis to adopt better strategies to fight crime. More specifically, predictive policing uses computer analysis of data to make predictions about future crimes to improve situational awareness and develop strategies that foster more effective and efficient policing⁹⁷. Considering the relevant studies on the topic, no single definition can be found⁹⁸.

⁹¹ Kent Police, *PredPol operational review - initial findings*. Restricted. Available at: <https://www.statewatch.org/media/documents/docbin/uk-2013-11-kent-police-pp-report.pdf>.

⁹² Kent Police, *Pred Pol operational review*. Restricted. Available at: <https://www.statewatch.org/media/documents/docbin/uk-2014-kent-police-predpol-op-review.pdf>.

⁹³ *Kent Police stop using crime predicting software*. Article published by the BBC, 26 November 2018. Available at: [Kent Police stop using crime predicting software](#).

⁹⁴ The idea of identifying three different phases in predictive policing can be found in A.G. FERGUSON, *Policing Predictive Policing*, in *Washington University Law Review*, 94(5), 2017, 1109-1189.

⁹⁵ See FERGUSON, *Policing Predictive Policing*, cited, 1126-1127. The reasons why the original focus was on property crimes may be identified in the following: first, property crimes normally generate a sense of insecurity in the community, even if they are not the most serious types of crime; second, they tend to be reported, therefore the police generally have a clear picture of this type of crime in their own territorial context; third, they are the most predictable, as criminological theories suggest (rational choice theory and the theory on routine activity).

⁹⁶ PredPol was mainly focused on property crimes, but it later evolved incorporating also gun-related violence.

⁹⁷ Similarly, EGBERT, LEESE, *Criminal Futures*, cited, 26. W.L. PERRY, B. MCINNIS, C.C. PRICE, S.C. SMIT, J.S. HOLLYWOOD, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*, RAND Corporation, 2013, 2.

⁹⁸ This is evident also from the analysis of Uchida. C.D. UCHIDA, *Predictive Policing*, in G. BRUINSMA, D. WEISBURD (Eds.), *Encyclopedia of Criminology and Criminal Justice*, Springer, 2014, 3862-3871. Available at: https://doi.org/10.1007/978-1-4614-5690-2_260.

CHAPTER 2

Author	Definition
William Bratton (2009)	“Predictive policing connects technology, management practices, real-time data analysis, problem-solving and information-led policing to lead to results—crime reduction, efficient police agencies, and modern and innovative policing” ⁹⁹
John Morgan (2009)	“Predictive policing refers to any policing strategy or tactic that develops and uses information and advanced analysis to inform forward-thinking crime prevention” ¹⁰⁰
Beth Pearsall (2010)	“Predictive policing, in essence, is taking data from disparate sources, analysing them and then using results to anticipate, prevent and respond more effectively to future crime” ¹⁰¹
Andrew Guthrie Ferguson (2012)	“Predictive policing has become a generic term for any crime fighting approach that includes a reliance on information technology (usually crime mapping data and analysis), criminology theory, predictive algorithms, and the use of this data to improve crime suppression on the streets” ¹⁰²
Walter L. Perry et al. (2013)	“Predictive policing is the application of analytical techniques – particularly quantitative techniques – to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions” ¹⁰³
Jerry Ratcliffe (2019)	Predictive policing means “the use of historical data to create a forecast of areas of criminality or crime hot spots, or high-risk offender characteristic profiles that will be one component of police resource allocation decisions. The resources will be allocated with the expectation that, with targeted deployment, criminal activity can be prevented, reduced or disrupted” ¹⁰⁴
Simon Egbert and Matthias Leese (2021)	Predictive policing is described as “the proactive use of algorithmically mediated data analysis for the purpose of finding patterns in datasets, based on which risk estimates are produced for either individuals or locations and are operationalised in the form of targeted prevention measures” ¹⁰⁵

⁹⁹ W. BRATTON, J. MORGAN, S. MALINOWSKI, *Fighting Crime in the Information Age: The Promise of Predictive Policing*. LAPD Research Paper, 2009, 3. Available at: <https://publicintelligence.net/lapd-research-paper-fighting-crime-in-the-information-age-the-promise-of-predictive-policing/>.

¹⁰⁰ C.D. UCHIDA, *A National Discussion on Predictive Policing: Defining Our Terms and Mapping Successful Implementation Strategies*, 18-19 November 2009, 1 reporting the definition given by the PhD student John Morgan.

¹⁰¹ PEARSALL, *Predictive Policing: The Future of Law Enforcement?*, cited, 1.

¹⁰² A.G. FERGUSON, *Predictive Policing and Reasonable Suspicion*, in *Emory Law Journal*, 62(2), 2012, 265.

¹⁰³ PERRY ET AL., *Predictive Policing*, cited, 1.

¹⁰⁴ J. RATCLIFFE, *Advocate: Predictive Policing*, in D. WEISBURD, A.A. BRAGA (Eds.), *Police Innovation. Contrasting Perspectives*, Cambridge University Press, 2019, 349.

¹⁰⁵ EGBERT, LEESE, *Criminal Futures*, cited, 19.

Among the several definitions, some cross-cutting themes can be identified. First, predictive policing has its roots in criminological theories, having criminology long been studying criminal activities and their relationship with space and time¹⁰⁶. Second, predictive policing is mainly based on data, such as data about past crimes, the local environment or other pertinent data (e.g., publicly accessible datasets on businesses, demographic data or social network data). Third, it relies on information and communication technologies (ICTs) to collect, store and analyse this data. Over the years, technological advances have made analytical tools increasingly complex and capable of operating on ever larger data sets, moving from simple heuristics to sophisticated mathematical techniques (e.g., machine learning), and of elaborating ever better predictions¹⁰⁷.

It is noteworthy, that elaborated predictions must go hand in hand with policing best-practice: once the police unit has a prediction of where and when crime is likely to occur, it must deploy resources in the best way, based on police officers' knowledge, skills and experience, to effectively have an impact on crime. Indeed, an algorithmic prediction, commonly expressed as a predictive score, quantifies the estimated probability that a given instance possesses a particular target property or that the target outcome is true¹⁰⁸. Predictions are just one component of the decision-making process, which is considered a broader process that ideally involves human judgment, contextual understanding, and ethical considerations¹⁰⁹. Therefore, algorithmic prediction – e.g., hotspot mapping or individual risk scores – must be interpreted by a human operator who shall then decide how to act based on the prediction, such as whether to act on that prediction, how to allocate resources, and how to balance competing priorities¹¹⁰.

Predictive policing is thus a complex socio-technical process, a “*process that is embedded within a set of complex sociotechnical relations*”¹¹¹: it begins with the collection of data (data collection) and continues with the analysis of this data by means of specific analytical techniques (data analysis); the results of the data analysis are then used by the police department to inform its subsequent actions

¹⁰⁶ D. WEISBURD, *The Law Of Crime Concentration and the Criminology Of Place in Criminology*, 53(2), 2015, 133–157.

¹⁰⁷ The reliance on predictions and predictive analytics makes some authors describe predictive policing as a «prediction-based police work». See J. SPRENGER, D. BRODOWSKI, ‘*Predictive policing*’, ‘*predictive justice*’, and the use of ‘*artificial intelligence*’ in the administration of criminal justice in Germany, Association internationale de droit pénal, 2023, 5. Available at: <https://penal.org/sites/default/files/files/A-02-23.pdf>.

¹⁰⁸ For an interesting discussion on predictive scoring and individuals see D.K. CITRON, F. PASQUALE, *The scored society: due process for automated predictions*, in 89 *Washington Law Review*, 1, 2014, 1-34.

¹⁰⁹ An interesting analysis on this topic can be found in F. LAGIOIA, R. ROVATTI, G. SARTOR, *Algorithmic fairness through group parities? The case of COMPAS-SAPMOC*, in *AI & SOCIETY*, 2022.

¹¹⁰ «Decision-making is not limited to predicting but also requires specifying the goals to be pursued, identifying the applicable ethical or legal constraints, evaluating the predicted consequences of alternative courses of action, and selecting the action best suited to goals and constraints». *Ibidem*, 3.

¹¹¹ EGBERT, LEESE, *Criminal Futures*, cited, 10. More precisely, the authors affirm that «predictive policing consists not only of software, algorithms and data sources» but hingers «on how police departments incorporate algorithmic crime analysis into their organizational structures and occupational cultures and how they manage to turn them into operational measures». *Ibidem*, 3.

(operational phase), the implementation of which results in an altered environment that leads to the collection of new data for analysis, triggering the process described again¹¹². Walter L. Perry et al. describe this process as a “prediction-led policing business process” (*Figure 8*)¹¹³ and Simon Egbert and Matthias Leese as a “chain of translation” (*Figure 9*)¹¹⁴, showing that the use of predictive analytics cannot be considered in isolation, but must be seen as part of the larger organisational structure of the police department.



Figure 8: Prediction-led policing business process. Source: Walter L. Perry et al, cited, 11

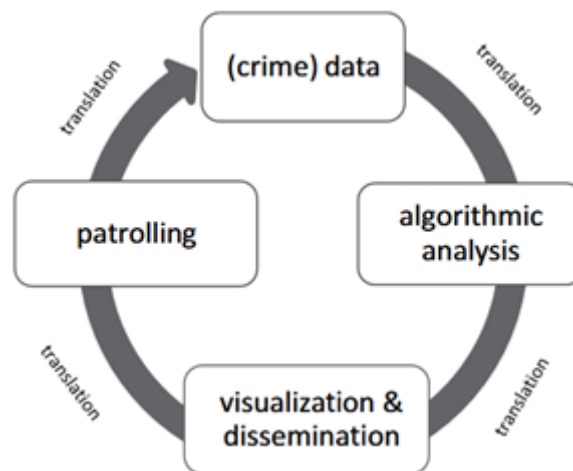


Figure 9: Chain of translation. Source: S. Egbert, M. Leese, *Criminal Futures*, cited, 4

¹¹² For a description of this process see also L.B. MOSES, J. CHAN, *Algorithmic prediction in policing: assumptions, evaluation, and accountability*, in *Policing and Society*, 28(7), 2018, 806–822.

¹¹³ PERRY ET AL., *Predictive policing*, cited, 11.

¹¹⁴ EGBERT, LEESE, *Criminal Futures*, cited, 4.

Simplifying the outlined process, it is possible to identify two different phases: i) the predictive phase and ii) the policing phase. The former was previously characterised by handcrafted formulas¹¹⁵ but now relies mainly on complex computer-implementable algorithms that can search and make matches across multiple databases¹¹⁶. By contrast, the latter focuses generally on more traditional forms of policing and operational measures¹¹⁷. To further investigate the topic, diverse predictive policing approaches are presented. Differentiating between approaches within predictive policing is relevant for this Thesis as the AI Act, as I will analyse in Chapter 3, regulates them differently.

1.2.1. Place-based predictive policing

Place-based approaches to predictive policing are concerned with the identification of geographic areas that are most likely to be prone to crime in each timeframe. Predictive policing software thus predicts the level of risk that a crime will be associated with a specific area in a specific time¹¹⁸. From a practical point of view, the data collected by the police is imported into the computer, the software runs on the data and creates a graphic visualisation of the areas at greatest risk of crime. The map is then used by police officers exercising their discretion.

Within place-based predictive policing, several methods and approaches can be identified. The aim of this paragraph is not to present all existing methods or to provide technical details on them, rather it is to emphasise the fact that there is a wide range of ways in which crime can be predicted, with considerable variation in terms of datasets, both in terms of quantity and type of data sources, and methods to be used, some of which are based on rule-based algorithms, others on adaptive algorithms. Consequently, different concerns may arise between different place-based predictive policing software.

One method used in place-based predictive policing is hot spot analysis. This method predicts areas (so-called “hot spots”) in which there is a greater likelihood of crime to occur. The basic idea is that

¹¹⁵ It was the case of Ernest Burgess that in the 1920s examined data from thousands of prisoners in order to deduce whether there was a pattern that does permit to identify those who would violate the condition of their parole. M. HAKEEM, *The Validity of the Burgess Method of Parole Prediction*, in *American Journal of Sociology*, 53(5), 1948, 376-386.

¹¹⁶ A detailed technical and legal discussion can be found in R.A. BERK, *Artificial Intelligence, Predictive Policing, and Risk Assessment for Law Enforcement*, in *Annual Review of Criminology*, 4, 2021, 209-237.

¹¹⁷ Crime problems are complex and therefore require complex responses, and traditional, localized policing practices are not always the best solution. Rather, crime problems would require a multi-agency approach. If the problem is with a place, for example, the solution is not simply deploying more police officers there but to improve the quality of housing, create better outdoor places, etc. PERRY ET AL., *Predictive Policing*, cited, 14 distinguishes between generic interventions and crime-specific and problem-specific interventions.

¹¹⁸ Perry et al. address the «first myth» of predictive policing, namely that «The computer actually knows the future». As they point out «these algorithms predict the risk of future events, not the events themselves. The computer, as a tool, can dramatically simplify the search for patterns, but all these techniques are extrapolations from the past in one way or another. In addition, predictions are only as good as the underlying data used to make them». *Ibidem*, Summary, xix.

crime is not evenly distributed in the urban context, as it is possible to identify areas with higher or lower crime rates, and the occurrence of crime tends to be stable over time¹¹⁹. Predictions in hot spot analysis are based on historical crime data and diverse techniques can be used to make predictions, such as grid mapping, covering ellipses, kernel density estimation or simple heuristic methods¹²⁰. The main advantages of hot spot analysis are the small amount of data required to make predictions and generally the non-complex techniques for analysing it; however, they lack the sophistication of other approaches.

A common method used is the near-repeat approach, which assumes that certain types of crime are likely to be followed by similar crimes occurring nearby in time and space. The near-repeat approach thus seeks to measure the time elapsed since the previous crime occurred and to derive useful information from this to make predictions about future crimes. It has its roots in George O. Mohler's idea that certain types of crime (e.g. domestic burglaries, gang shootings) present similar characteristics with earthquakes, envisioning that self-exciting point processes from seismology can be used for the purpose of crime prediction¹²¹. Like hot spot analysis, this approach requires few data points as inputs, namely crime data, and the model and algorithms are not particularly complex. Examples of software that operationalise the near-repeat approach are PreMap (Predictive Policing Mobile Analytics for Police) and PredPol.

Risk terrain modelling is another approach followed in place-based predictive policing¹²². The main assumption is that crime can be forecasted based on different data points, such as geographic features (e.g., metro stations, parks, bars, specific roads, shops, etc.), that are normally associated with crime. Based on this assumption, risk terrain techniques attempt to make predictions on crime based on the distance between a given location and the risk-inducing features. In recent years, work has begun to explore how to dynamically model crime risk in relation to the movement of people across the city, considering other data, such as traffic data, public transport data, weather data, etc. The advantage of

¹¹⁹ PERRY ET AL., *Predictive Policing*, cited, 19. EGERT, LEESE, *Criminal Futures*, cited, 31. L. SHERMAN, P. GARTIN, M. BUERGER, *Hot spots of predatory crime: Routine activities and the criminology of place*, in *Criminology*, 27(1), 1989, 27–56. A.A. BRAGA, *Hot Spots Policing: Theoretical Perspectives, Scientific Evidence, and Proper Implementation*, in B. TEASDALE, M. BRADLEY (Eds.), *Preventing Crime and Violence*, Springer, 2017, 269–279. K. BOWERS, S.D. JOHNSON, K. PEASE, *Prospective Hot-Spotting: The Future of Crime Mapping?* in *British Journal of Criminology*, 2004, 44(5), 641–658.

¹²⁰ For a discussion of such methods see PERRY ET AL., *Predictive Policing*, cited.

¹²¹ G.O. MOHLER, M.B. SHORT, P.J. BRANTINGHAM, F.P. SCHOENBERG, G.E. TITA, *Self-Exciting Point Process Modeling of Crime*, in *Journal of the American Statistical Association*, 106(493), 2011, 100–108. Other studies confirmed Mohler's hypothesis. M. TOWNSLEY, R. HOMEL, J. CHASELING, *Infectious burglaries. A test of the near repeat hypothesis*, in *British Journal of Criminology*, 43(3), 2003, 615–633. M. TOWNSLEY, R. HOMEL, J. CHASELING, *Repeat Burglary Victimization: Spatial and Temporal Patterns*, in *Australian and New Zealand Journal of Criminology*, 33(1), 2000, 37–63.

¹²² For the following description reference is made to J.M. CAPLAN, L.W. KENNEDY, *Risk terrain modelling manual: theoretical framework and technical steps of spatial risk assessment for crime analysis*, Rutgers Center on Public Security, Newark, NJ, 2010.

risk terrain techniques is that they do not make predictions based on past crime data alone, but also consider geographical attributes, the socio-economic composition of the area and other characteristics, allowing the software to predict hot spots in new and different areas, whose characteristics are like areas where previous crimes have been committed.

The identification of areas of increased crime risk can also be obtained through the building of mathematical models for generating predictions based on data¹²³. Techniques used to achieve this result are generally referred to as data mining, which can be understood as the process of discovering models or patterns in large volumes of data¹²⁴. More specifically, data mining uses different techniques to automatically or semiautomatically seek models or patterns in data and use them to explain the data and make predictions from it. In recent years, the opportunities offered by data mining have grown, thanks to the increase in data and in computing capabilities and in the improvement of algorithms that can perform the search¹²⁵. Different techniques can be used for data mining, varying in complexity: at one extreme, there are very simple methods which use a few variables and build a mathematical model with simple relationships between the variables (if-then statements); at the other extreme, there are methods that considers many different variables and combine them through complex relationships to generate forecasts. Three families of data mining techniques are of primary interest when predictive policing is involved, most of them developed within the field of machine learning: regression, clustering and classification¹²⁶.

Furthermore, several approaches can be used together to develop software that generates a unified crime prediction. HunchLab 2.0 represents an interesting example. As described by Azavea¹²⁷, the developer of HunchLab, the software has been enhanced with more predictive capabilities than the previous version (HunchLab 1.0) through the incorporation of variables – such as temporal and weather patterns, risk terrain modelling features, socio-economic factors, historic crime data, near-repeat patterns – into a single prediction of crime risk. The system then uses machine learning approaches to learn which patterns are relevant for each type of crime and generate predictions about the probability of a particular crime occurring in certain locations in each timeframe¹²⁸. HunchLab

¹²³ PERRY ET AL., *Predictive Policing*, cited, 29.

¹²⁴ WITTEN, FRANK, *Data Mining*, cited, 5. Data mining can be defined as the extraction of knowledge from data, via technologies that incorporate principles of data science.

¹²⁵ For a technical discussion see A. BOGOMOLOV, B. LEPRI, J. STAIANO, N. OLIVER, F. PIANESI, A. PENTLAND, *Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data*, arXiv:1409.2983v1, 10 September 2014.

¹²⁶ M.J. ZAKI, M. WAGNER, *Data Mining and Machine Learning: Fundamental Concepts and Algorithms*, Cambridge University Press, 2020.

¹²⁷ Azavea, *HunchLab: Under the Hood*, 2015. Available at: <https://blog.pilpul.me/files/2015/09/HunchLab-Under-the-Hood.pdf>.

¹²⁸ Azavea points out that «in some ways, the model building process in HunchLab mimics the thought process of an experienced analyst». *Ibidem*, 16.

thus uses much more data and more complex techniques to make predictions than PredPol or similar software.

Considering the above, the main differences between the predictive policing software of the 1990s and today's can be summarised as follows: the amount and diverse types of data which can be used by the software (such as criminal data, social media data, traffic data, weather data); the techniques used to analyse this data, which now includes also more complex machine learning algorithms that allow the machine to learn and create the model from given input data; and the combination of different approaches in order to obtain better predictions¹²⁹. Not all models generate equally accurate predictions or are equally efficient¹³⁰.

1.2.2. Person-based predictive policing

From a legal perspective, person-based predictive policing raises more concerns than place-based predictive policing, as the software makes predictions about the risk of an individual committing a crime or being the victim of a crime¹³¹. As in place-based predictive policing, in relation to person-based predictive policing different approaches can be identified. However, all these approaches bear on the assumption that there are certain characteristics of the individual and his or her social network that can be used as indicators to predict the individual's future actions¹³².

Two main approaches can be identified, namely risk profiling and social network analysis. The first approach resembles the type of software used in the U.S. criminal justice system, such as COMPAS, in that it compares individual characteristics with those considered risk factors of known offenders. Based on this comparison, the software makes a prediction about the risk of the subject committing a crime or being the victim of a crime. Therefore, a risk profile is drawn up at an individual level and it can be used by the operator to make decisions on subsequent operations. One example of person-based predictive policing is the Strategic Subject Algorithm. The algorithm was created by the Illinois Institute of Technology and used by the Chicago Police Department for the creation of a risk assessment score reflecting the likelihood of an individual to be involved in a shooting incident, either

¹²⁹ The combination of different techniques might lead to more accurate outcomes, as recognised by G.M. CAMPEDELLI, *Machine Learning for Criminology and Crime Research*, cited, 56-57.

¹³⁰ From a general point of view, models are not the same and can present differences in terms of accuracy and efficiency. G. SARTOR, *L'informatica giuridica e le tecnologie dell'informazione*, cited, 88-96.

¹³¹ This is the reason why some scholars refer to person-based predictive policing as «predictive profiling». H. LAMMERANT, P. DE HERT, *Predictive profiling and its legal limits: Effectiveness gone forever*, in B. VAN DER SLOOT, D. BROEDERS, E. SCHRIJVERS (Eds.), *Exploring the boundaries of big data*, Amsterdam University Press, 2016, 145-173. L.M. SOMMERER, *Geospatial Predictive Policing – Research Outlook & A Call For Legal Debate*, in *Neue Kriminalpolitik*, 29(2), 2017, 149.

¹³² EGBERT, LEESE, *Criminal Futures*, cited, 28.

as a victim or an offender. The scores ranged from “extremely low risk” to “extremely high risk”, giving rise to a rank-ordered list, known as the Strategic Subject List (SSL). The score is generated based on data which includes attributes of a person’s criminal record, the gravity of his/her criminal history and the increase or decrease in his/her criminal activities¹³³. Based on this list, people identified as at risk of gun-related violence are informed that they are under the watchful eye of the police in order to deter them committing a crime. The list is then used to guide subsequent police interventions. In the United States there are many other examples based on the same approach¹³⁴.

Social network analysis, on the other hand, focuses on the concept of network¹³⁵. In this case, dangerousness is inferred based on the subject’s social contacts (such as friends, neighbours, relatives, colleagues), on the assumption that links with people who have been perpetrators or victims of a (violent) crime in the past may be an indicator for the commission of a similar offence in the future¹³⁶. This approach assumes that negative social networks may encourage deviant future behaviours (*i.e.* criminal activities). In the United States, to encourage local police departments to innovate their work, a national initiative, called Smart Policing Initiative (SPI), was adopted by the Department of Justice’s Bureau of Justice Assistance (BJA)¹³⁷. To identify probable offenders in Kansas City, for example, a social network analysis was conducted by the local police department funded through the SPI. The analysis started with the initial offender, going through his associates and finally analysing the associates of the initial offender’s associates. The presumed associates were then contacted by the police and informed that they had been identified as a “probable cause of violence” in the city, they were informed about available social services and advised that they would face harsher punishment if they committed a crime¹³⁸.

Another example of person-based predictive policing funded through the SPI is Operation LASER (short for Los Angeles’ Strategic Extraction and Restoration Program), started in Los Angeles in

¹³³ Chicago Police Department, Special Order S10-05, § V., October 6, 2015. Available at: <http://directives.chicagopolice.org/directive>.

¹³⁴ Similar examples are the VOID (short for Violent Offender Identification Directive), the KC No VA (short for Kansas City No Violence Alliance) and the ODARA (short for Ontario Domestic Assault Risk Assessment). For a brief description of these programmes see M. HAMILTON, *Predictive policing through risk assessment*, in J.L.M. MCDANIEL, K.G. PEASE, *Predictive Policing and Artificial Intelligence*, Routledge, 2021, 62-63.

¹³⁵ R. VAN DER HULST, *Introduction to Social Network Analysis (SNA) as an investigative tool*, in *Trends in Organized Crime*, 12, 2009, 101-121.

¹³⁶ EGBERT, LEESE, *Criminal Futures*, cited, 29. For a complete and detailed analysis of social network analysis see M.A. TAYEBI, U. GLÄSSER, *Social Network Analysis in Predictive Policing*, Springer, 2016.

¹³⁷ A.A. BRAGA, D.W. WEBSTER, M.D. WHITE, H. SAIZOW, *SMART Approaches to Reducing Gun Violence: Smart Policing Initiative Spotlight on Evidence-Based Strategies and Impacts*, 2014, 13. Available at: <https://www.ojp.gov/ncjrs/virtual-library/abstracts/smart-approaches-reducing-gun-violence-smart-policing-initiative>.

¹³⁸ *Ibidem*, 13.

2012¹³⁹. The aim of the strategy is to identify places where gun-related crimes are likely to occur¹⁴⁰ and to target individuals at risk of committing violent crimes. To briefly describe how the offender-based strategy works, different phases can be identified: first, intelligence is gathered from daily patrols (e.g., data on crime reports, arrest reports, field interview cards); second, this information is used to create a list of “Chronic Offenders”; third, the police use the software acquired from Palantir Technologies Incorporation (hereinafter Palantir) to generate “Chronic Offender Bulletins” which are information documents that contain the name and other personal information of the offender (e.g., address, driver’s licence number, arrest history, physical characteristics, police contacts, etc.); fourth, the Bulletins are provided to the police. Since the subjects listed in the Bulletins are not necessarily under criminal investigation (they are not necessarily wanted for a crime, or they have no pending arrest warrants), the police cannot arrest or detain the individual based on the sole information contained in the Bulletin. Police officers are thus requested to gather intelligence about them during their daily patrol activity and fill out a field interview card. To identify offenders who deserve more attention from the police, a “point system” is used: offenders are given a point value (e.g., five points for a violent crime, five points for a known gang affiliation, one point for a quality police contact, etc.)¹⁴¹ that eventually create a risk score, and they are rank accordingly.

This is a clear example of advanced analytics used to collect huge amounts of data, gather intelligence on a specific subject and create a detailed profile of the targeted individual based on his/her personal information, such as personal characteristics, prior criminal history and contacts with the police, which is defined individual risk profiling. As Sarah Brayne stated, reporting the words of a Palantir software engineer working for LAPD, it is surprising “*how little pieces of data that might seem innocuous at the time of the collection can eventually be pulled together to create useful intelligence*”¹⁴². Already from this brief description, some concerns can be highlighted, which will be analysed in more detail later in the Chapter. First, the collection and processing of huge amounts of personal data through software that enables data analysis is extremely critical from the perspective of data protection legislation. Second, the “risk scoring system” can easily become discriminatory, due to a feedback-loop effect: police officers are called upon to focus their attention on individuals with higher risk scores, the increased surveillance of these subjects is likely to lead to the detection of misbehaviour

¹³⁹ Operation LASER was described in detail by Sarah Brayne. See BRAYNE, *Predict and Surveil*, cited, 60-70. C.D. UCHIDA, M. SWATT, D. GAMERO, J. LOPEZ, E. SALAZAR, E. KING, R. MAXEY, N. ONG, D. WAGNER, M.D. WHITE, *Los Angeles, California Smart Policing Initiative: Reducing Gun-Related Violence Through Operation LASER*, 2012. Available at: <https://www.ojp.gov/ncjrs/virtual-library/abstracts/los-angeles-california-smart-policing-initiative-reducing-gun>.

¹⁴⁰ In this sense, Operation LASER resembles hotspot analysis in that it creates heat maps for weapons-related crimes. However, this programme goes a step further by also identifying individuals at risk of committing crimes.

¹⁴¹ UCHIDA ET AL., *Los Angeles, California Smart Policing Initiative*, cited, 8.

¹⁴² BRAYNE, *Predict and Surveil*, cited, 64.

with possible increased contact of them with the police, which in turn leads to an increase in the risk score, increasing the likelihood that the individual will be stopped again in the future. Third, discrimination can also result from bias in the data with which the software is fed. Fourth, private companies are acquiring an ever-higher role in the law enforcement sector, being those who develop the software (which is generally proprietary) and store and analyse police data or collect data they then sell to the police. Palantir, for example, is one of the biggest private vendors of predictive policing software with which the U.S. police is working¹⁴³, but it is just one of the many¹⁴⁴.

1.3. Predictive policing in Europe

More recently, police authorities in European countries have started to develop or purchase predictive policing software and implement them in their organisations¹⁴⁵. However, the use of these algorithmic tools is still at an early stage and characterised by departmental diversity. So far, only a few studies have mapped predictive policing practices in Europe in recent years¹⁴⁶, showing that the police departments are increasingly interested in data and its analysis to inform their strategic decisions and operational measures.

In this section, I will present some cases of predictive policing, which I found relevant as they allow me to investigate more general aspects related to the use of AI in law enforcement, namely bias and discrimination, transparency and explainability, accountability and human oversight. As a preliminary remark, it should be emphasised that European police forces are very diverse, the police function being closely linked to the peculiarities of each State. The aim of this empirical research is to describe these practices objectively, highlighting the relevant aspects that allow me to reach more general considerations on the issues connected to the use of these systems in the European context.

¹⁴³ Palantir has started a partnership with the City of New Orleans in the NOLA For Life Project, which is aimed at reducing homicides in New Orleans. See City of New Orleans, *NOLA For Life. Comprehensive Murder Reduction Strategy*, 2013. Available at: https://www.nola.gov/nola/media/Mayor-s-Office/Files/NOLA%20FOR%20LIFE/NOLA-FOR-LIFE-Sept2013-book-pdf_1.pdf. Palantir, *Nola Murder Reduction. Technology to power data-driven public health strategies*, 2014. Available at: <https://www.documentcloud.org/documents/5017843-NOLA-Murder-Reduction-White-Paper/>.

¹⁴⁴ IBM, PredPol, Aveza are other examples of private companies working for U.S. law enforcement authorities.

¹⁴⁵ See Algorithm Watch, *Automating Society Report 2020*, October 2020. Available at: <https://automatingsociety.algorithmwatch.org>.

¹⁴⁶ One of the most interesting works is EGBERT, LEESE, *Criminal Futures*, cited. Although it is focused on German-speaking countries general insights on predictive policing can be identified. F. JANSEN, *Data-driven policing*, cited considers more countries: United Kingdom, the Netherlands and Belgium. M. OSWALD, J. GRACE, *Intelligence, policing and the use of algorithmic analysis: a freedom of information-based study*, in *Journal of Information Rights, Policy & Practice*, 1(1), 2016, 1-14. An analysis that look both at the European and at the U.S. context is E. PIETROCARLO, *Predictive policing: criticità e prospettive dei sistemi di identificazione dei potenziali criminali*, in *Diritto Penale Contemporaneo*, 2, 2023, 114-158.

Predictive policing is currently widespread throughout Europe, especially in the northern countries where investment in advanced technologies is greater and the introduction of technologies in the public sector is more advanced¹⁴⁷. United Kingdom, Germany and the Netherlands are also experimenting with predictive policing. In European countries both place-based and person-based predictive policing software is used.

As far as the latter is concerned, in the United Kingdom, the Durham Constabulary in partnership with the University of Cambridge, developed the Harm Assessment Risk Tool (HART)¹⁴⁸. The idea on which HART is based is that offenders should be treated differently, depending on their dangerousness, to reduce reoffending and prevent further crimes¹⁴⁹. HART is based on a machine learning algorithm (*i.e.* random forest¹⁵⁰), based on multiple predictor variables¹⁵¹, to predict offender dangerousness and classify offenders into different risk groups (high, moderate, low). Offender dangerousness is evaluated based on the likelihood of the offender committing a “serious offence” (high), a “non-serious offence” (moderate) or “no offence” (low). The software is used after the offender’s arrest and before his/her entrance in the Criminal Justice System and the prediction is used to support the subsequent decision-making process on custody disposal. Those arrestees forecast as “moderate risk” are eligible

¹⁴⁷ The reference goes to Norway, Sweden, Denmark. As far as these countries are concerned see H.O.I. GUNDHUS, P.E. SKJEVRÅK, C.T. WATHNE, *We Will Always Be Better Than a Spreadsheet. Intelligence Logic and Crime Prevention in Practice*, in *European Journal of Policing Studies*, 6(1), 2023, 27-49. H.O.I. GUNDHUS, C.T. WATHNE, *Resistance to platformization: Palantir in the Norwegian police*, in *Information, Communication & Society*, 27(13), 2024, 2457-2475. Available at: <https://doi.org/10.1080/1369118X.2024.2325533>. MASSO ET AL., *Citizens’ perspectives on platformisation of police work*: cited. V. GALIS, B. KARLSSON, *A world of Palantir – ontological politics in the Danish police’s POL-INTEL*, in *Information, Communication & Society*, 27(13), 2024, 2438-2456.

¹⁴⁸ M. OSWALD, J. GRACE, S. URWIN, G.C. BARNES, *Algorithmic risk assessment policing models: lessons from the Durham HART model and ‘Experimental’ proportionality*, in *Information & Communications Technology Law*, 27(2), 2018, 224. G. BARNES, *Focusing Police Resources: Algorithmic Forecasting in Durham*. Paper presented to the 9th International Conference on Evidence-Based Policing, Cambridge, United Kingdom, 16 July 2016. Interesting research on the programme HART and its accuracy was conducted by S. URWIN, *Algorithmic forecasting of offender dangerousness for police custody officers: an assessment of accuracy for the Durham Constabulary Model*, Master Thesis in Applied Criminology and Police Management, University of Cambridge, 2016. The Thesis was written under the supervision of Geoffrey Barnes and research was conducted at the University of Cambridge and the Durham Constabulary.

¹⁴⁹ An interesting analysis on the topic was conducted by L.W. SHERMAN, P.W. NEYROUD, *Offender-Desistance Policing and the Sword of Damocles*, Civitas: Institute for the Study of Civil Society, London, 2012. L.W. SHERMAN, *Offender-Desistance Policing (ODP). Less Prison and More Evidence in Rehabilitating Offenders*, in T. BLIESENER, A. BEELMANN, M. STEMMLER (Eds.), *Antisocial behavior and crime: Contributions of developmental and evaluation research to prevention and intervention*, Hogrefe Publishing, 2012, 199-218.

¹⁵⁰ «Random forest modeling techniques represent an improvement over the methodologies of traditional risk prediction instruments. Random forests allow for the inclusion of a large number of predictors, the use of a variety of different data sources, the expansion of assessments beyond binary outcomes, and taking the costs of different types of forecasting errors into account when constructing a new model». G. BARNES, J.M. HYATT, *Classifying Adult Probationers by Forecasting Future Offending*, Final Technical Report, March 2012, 3. Available at: <https://www.ncjrs.gov/pdffiles1/nij/grants/238082.pdf>.

¹⁵¹ HART considers thirty-four predictor variables. See OSWALD, GRACE, URWIN, BARNES, *Algorithmic risk assessment policing models*, cited, 228.

for the Constabulary's Checkpoint programme¹⁵². From this first example, it is evident that predictive policing might be something more than just a simple technical tool used by police forces to make their work more effective and efficient. Rather, it is part of a social process that aims, at least in the case of HART, to address the root causes of crime and to pay attention to the welfare of the community¹⁵³. From a more technical point of view, the HART model uses thirty-four different predictors to arrive at a forecast, most of them focused upon the criminal history of the offender, and more than five hundred classification and regression trees combined in the forecasting model¹⁵⁴. In addition to HART, the United Kingdom has opened up to other programmes, both commercial and non-commercial¹⁵⁵. Most predictive policing practices in Europe, however, focus on identifying when and where crime is likely to be committed. In the Netherlands, the Dutch Police started using the Crime Anticipation System (CAS), a predictive policing system developed in house¹⁵⁶. It targets high impact crimes (e.g. domestic burglaries, robberies and mugging) and other crimes (e.g. pickpocketing, car burglaries, violent crimes, commercial burglaries and bicycle theft) displaying areas of higher crime risk through the construction of heat-maps¹⁵⁷. It uses local and current data derived from three data sources¹⁵⁸: i) socio-economic data (from the Central Bureau of Statistics), ii) location specific characteristics (number and kind of companies in the area, distance to closest known offender) and iii) crime history (number of crimes in different time periods)¹⁵⁹. Furthermore, data from the Municipal Administration (e.g., streets, addresses) are used as structure for the map.

¹⁵² The Constabulary's Checkpoint programme was described as a «*culture-changing initiative within Durham Constabulary*». *Ibidem*, 227. Checkpoint programme webpage, Durham Constabulary <https://www.durham.police.uk/Information-and-advice/Pages/Checkpoint.aspx>.

¹⁵³ OSWALD, GRACE, URWIN, BARNES, *Algorithmic risk assessment policing models*, cited, 227. However, this is not always the case. In most cases, predictive policing software is said to be directed at immediately managing unwanted behaviours that addressing the root causes of crime. See M. ANDREJEVIC, *Automated Media*, Routledge, 2020, 79–80. Similarly, JANSEN, *Data-driven policing: Negotiating the legitimacy of the police*, cited, 22.

¹⁵⁴ OSWALD, GRACE, URWIN, BARNES, *Algorithmic risk assessment policing models*, cited, 228. For a similar forecasting model in a domestic violence context, see R.A. BERK, S.B. SORENSON, G. BARNES, *Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions*, in *Journal of Empirical Legal Studies*, 2016, 13(1), 94–115.

¹⁵⁵ See H. COUCHMAN, *Policing by machines*, Report, Liberty, January 2019. Available at: [Report: Policing by machine - Liberty](#).

¹⁵⁶ For a description of big data and policing in the Netherlands see M. SCHUILENBURG, M. SOUDIIN, *Big data policing: The use of big data and algorithms by the Netherlands Police*, in *Policing: a journal of policy and practice*, 17, 2023, Article paad061.

¹⁵⁷ For a detailed analysis of CAS see S. OOSTERLOO, G. VAN SCHIE, *The Politics and Biases of the "Crime Anticipation System" of the Dutch Police*, Proceedings of the International Workshop on Bias in Information, Algorithms, and Systems, 2103, 2018, 30–41.

¹⁵⁸ The sources are: i) the Central Crime Database, ii) the Municipal Administration, and iii) the Central Bureau of Statistics of the Netherlands. *Ibidem*, 3.

¹⁵⁹ D. WILLEMS, *CAS: Crime Anticipation System. Predictive Policing in Amsterdam*. Available at: https://event.cwi.nl/mtw2014/media/files/Willems.%20Dick%20-%20CAS%20Crime%20anticipation%20system%20_%20predicting%20policing%20in%20Amsterdam.pdf.

PRECOBS (short for *Pre-Crime Observation System*) is a German-language place-based predictive policing software used in multiple departments in Germany, Switzerland and Austria since 2015/2016¹⁶⁰ and developed by the German manufacturer IfmPt (short for *Institut für musterbasierte Prognosetechnik GmbH*, which is now LogObject). The software is based on the near repeat approach¹⁶¹, like PredPol, and is specialised on residential burglaries. The main aim of PRECOBS is to apply effective preventive strategies by allocating police officers where there is a higher risk that burglaries will be committed, to prevent them and targeting the offender criminal type (the “professional serial burglar”). As Egbert and Leese, who deeply analysed the program¹⁶², point out the model and algorithms used by PRECOBS are not particularly complex and the results can be interpreted quite easily. Indeed, they stated that PRECOBS essentially automated previous manually performed crime analysis¹⁶³ using “*historical spatiotemporal distributions of crime as the basis for dynamic estimates of future risk*”¹⁶⁴.

In Germany, many other software used or in use were developed in-house by the respective police departments¹⁶⁵. While place-based approaches have been for the past years the prevalent type of predictive policing, recently the German police force went a step further and developed a few software based on person-based approaches. At the federal level, RADAR-iTE is an example of a software based on the risk assessment approach. Its aim is to assess the risk of individuals, already identified by the police, of committing Islamist-motivated terrorist acts and prioritise police interventions and allocate resources accordingly. The software classifies individual, either as “high-risk” or “moderate risk”, based on a form with standardised questions and answers which is filled out by police officers with information they had already gathered. Those individuals who are considered high-risk are kept under the watchful eye of the *Bundeskriminalamt* (Federal Criminal Police Office). The software was fully deployed at the national level in 2017 and then revised in 2019 in cooperation with the University

¹⁶⁰ SPRENGER, BRODOWSKI, ‘Predictive policing’, ‘predictive justice’, and the use of ‘artificial intelligence’ in the administration of criminal justice in Germany, cited, 7. As reported by the authors, Baden-Württemberg and Bavaria decided to stop its implementation for the reason that there was not enough available data for the software to work effectively.

¹⁶¹ D.A. BALOGH, *Near Repeat-Prediction mit PRECOBS bei der Stadtpolizei Zürich*, in *Kriminalistik*, 70(5), 2016, 335–341. D. GERSTNER, *Predictive policing in the context of residential burglary: An empirical illustration on the basis of a pilot project in Baden-Württemberg, German*, in *European Journal for Security Research*, 3, 2018, 115–138.

¹⁶² The reference goes to EGBERT, LEESE, *Criminal Futures*, cited.

¹⁶³ *Ibidem*, 8.

¹⁶⁴ *Ibidem*, 33.

¹⁶⁵ Among the predictive software it is possible to cite: KLB-operativ, KrimPro, PreMAP, SKALA.

of Konstanz¹⁶⁶. It is worth noting that, while it processes data automatically, the model used is not particularly complex¹⁶⁷.

Recently, another software used by some German Länder has attracted the attention of the community: hessenDATA, which has been used in Hessen since 2017, and is the result of the customisation of the Gotham platform sold by the U.S. private company Palantir. The software processes huge amount of data derived from heterogeneous sources, such as crime data, data from communication service providers, data from electronic devices, etc. These data from external sources are not integrated directly by the software but are retrieved by the police authorities on the basis of the power granted to them by law and only then are they integrated into the software¹⁶⁸. Since the main objective of hessenDATA is to identify matches, links and patterns between data, it resembles that software which is based on the social network analysis approach. However, it could be used also to carry out an individual risk assessment¹⁶⁹. It is noteworthy that the legal basis for hessenDATA was found in §25a of the Hessian Act on Public Security and Order (*Hessisches Gesetz über die öffentliche Sicherheit und Ordnung*, HSOG), which was specially introduced by the state legislature¹⁷⁰. After Hessen, other German Länder, namely Hamburg, North Rhine-Westphalia and Bavaria, showed their interest in the Palantir's software and some of them adopted a legal basis for its use¹⁷¹.

Germany is not the only European country doing business with Palantir. Norway, Sweden and Denmark have purchased software from the U.S. company, each with its own characteristics. In Denmark, for example, Palantir's software version is called POL-INTEL and is specifically

¹⁶⁶ See Federal Criminal Police Office, *RADAR (Regelbasierte Analyse potentiell destruktiver Täter zur Einschätzung des akuten Risikos)*. Available at: https://www.bka.de/DE/UnsereAufgaben/Deliktsbereiche/PMK/Radar/radar_node.html;jsessionid=9AB1BDE4A134C483F1820378A09EAF6A.live612#doc142872bodyText4.

¹⁶⁷ In SPRENGER, BRODOWSKI, 'Predictive policing', 'predictive justice', and the use of 'artificial intelligence' in the administration of criminal justice in Germany, cited, 10 the authors report that RADAR-iTE is based on Microsoft Excel. As the authors reports the legal basis for RADAR-iTE is §18(3) in connection with §18(1) No 4, §16 BKAG (Federal Criminal Police Office Act). While the provision does not mention the use of technology to process data, it refers «to the 'further processing of personal data' in case of indications that a person concerned is likely to commit a serious crime in the future». *Ibidem*, 10.

¹⁶⁸ SPRENGER, BRODOWSKI, 'Predictive policing', 'predictive justice', and the use of 'artificial intelligence' in the administration of criminal justice in Germany, cited, 13.

¹⁶⁹ *Ibidem*, 15.

¹⁷⁰ In 2023, the Bundersverfassungsgericht ruled on the question of the constitutionality of the legal bases on which the Länder Hessen and Hamburg had legitimised the use of the software, declaring them unconstitutional because they had violated the *Recht auf informationelle Selbstbestimmung* (right to informational self-determination). *Bundersverfassungsgericht*, I Senate (President Stephan Harbarth), February 16, 2023 – 1 BvR 1547/19, 1 BvR 2634/20. For an initial comment on the judgment see R. ORLANDI, *Uso poliziesco dell'intelligenza artificiale. L'insegnamento del Bundersverfassungsgericht*, in *Cassazione Penale*, 7/8, 2023, 2167-2180; M. MERLER, *La Corte costituzionale tedesca si pronuncia sul trattamento automatizzato dei dati in contesti di polizia*, in *Giornale di diritto amministrativo*, 5, 2023, 685-694.

¹⁷¹ Hamburg adopted Section 49 of the *Gesetz über die Datenverarbeitung der Polizei* (Hamburg Act on Data Processing by the Police – PolDVG) and North Rhine-Westphalia Section 23(6) of the *Polizeigesetz des Landes Nordrhein-Westfalen* (North Rhine-Westphalia Police Act – PolG NRW).

customised for the Danish police¹⁷². POL-INTEL is a platform for “*compiling, visualizing and analyzing data*”¹⁷³ which connects multiple police and non-police databases, such as the National Motor Vehicle registry, the National Photo registry, the police case management system (POLSAS) or international databases (e.g., Interpol database).

The use of Palantir’s software has been and is being criticised in all countries for several reasons. First, Palantir is a multi-billion-dollar private company whose main interest is to do business and make money. There is no room in its policy for considerations of protecting the rights and freedoms of individuals. Second, Palantir has strict connections with the U.S. politics¹⁷⁴, U.S. police departments¹⁷⁵ and, more importantly, with the U.S. Intelligence Service¹⁷⁶. This may pose a danger to the rights of individuals but could potentially endanger the very independence of European countries. Third, from a technical point of view, Palantir’s software allows for particularly advanced data analysis of huge quantities of data. Therefore, while the company presents itself as a creative producer of “*generalizable platforms for modeling the world and making decisions*”¹⁷⁷, the reality is that its products allow one to rapidly access thousands of data (personal and non-personal) across various databases to find out links, patterns and connections between the data which can be used for follow-up operations.

In Italy, there are currently no publicly available information on the acquisition of such software by police departments. Conversely, other predictive policing programmes have been developed over the years. Up until 2008, the Milan Police Department was the only one using predictive policing to combat crime¹⁷⁸. Specifically, software KeyCrime is focused on commercial robberies and analyses large datasets related to individual characteristics of robbers (e.g., gesture of the offender, physical characteristics, etc.) and individual criminal strategies collected from the official report of the police, victim reports and closed-circuit security cameras. Specifically, KeyCrime collects and analyse micro-level data related to robberies – which includes data about the crime, the observed perpetrator, the observed weapons, the mode of transportation used by the perpetrator¹⁷⁹ – that take place in the

¹⁷² See GALIS, KARLSSON, *A world of Palantir*, cited.

¹⁷³ *Ibidem*, 2439. The authors report the definition given by a POL-INTEL Project Owner during an interview conducted in September 2021.

¹⁷⁴ The reference goes to one of the founders of Palantir Technologies, Peter Andreas Thiel, who is a businessman but also an American politician.

¹⁷⁵ Among the clients, the NYPD and the LAPD are the «big guns». BRAYNE, *Predict and Surveil*, cited, 8.

¹⁷⁶ BRAYNE, *Predict and Surveil*, cited, 7. As reported by the author, among Palantir’s clients there are the Central Intelligence Agency (CIA) and the Federal Bureau of Investigation (FBI), the Immigration and Customs Enforcement (ICE) and the Department of Homeland Security (DHS).

¹⁷⁷ Palantir, *Direct Listing Letter*, August 25, 2020. Available at: <https://www.palantir.com/newsroom/letters/direct-listing-letter/>.

¹⁷⁸ For the following discussion see G. MASTROBUONI, *Crime is Terribly Revealing: Information Technology and Police Productivity*, in *The Review of Economic Studies*, 87(6), 2020, 2727–2753.

¹⁷⁹ *Ibidem*, 12.

municipality of Milan and helps the operator to find matches between robbers or groups of robbers characterised by the same *modus operandi*. Once the matches are established, the information is used to generate predictions based on statistical and criminological models. The potential future targets are then communicated to police patrols along with information related to the likely day, time, location of the future crime and a description of the *modus operandi* of the anonymous offender. KeyCrime can be described as a hybrid software, which does not fit completely into the place-based approach, but neither does it fit into the person-based approach. Indeed, it is based on the crime linking approach, the purpose of which is to anonymously profile the potential perpetrator of serial crimes (e.g. burglaries). The information obtained can be used both to solve past crimes and to predict future ones, on the assumption that some crimes (e.g. robberies or, more generally, street crimes) are characterised by certain regularities that make them more predictable.

Considering the experience of KeyCrime, the Italian Ministry of the Interior – Department of Public Security started developing the software Giove, a programme for the automated processing and analysis of data for the prevention and repression of crimes of high social impact (e.g., robberies, rape). Unfortunately, there is currently not much publicly available information on the Giove software¹⁸⁰. However, although it is based on the same crime linking approach as KeyCrime, aimed at identifying the same *modus operandi* in serial crimes, it is said to go a step further, implementing advanced AI to improve predictions¹⁸¹. However, as of 2023, the Central Anticrime Directorate confirmed it remained in the “feasibility study” phase and was not yet operational. Meanwhile, a parliamentary request has been sent from some parliamentarians to the Ministry of the Interior, asking for more information about the project¹⁸².

¹⁸⁰ Some insights can be found in journalistic articles. See I. CIMMARUSTI, B.L. MAZZEI, *Nome in codice Giove, ecco l'algoritmo della polizia per prevenire i reati*. Article published on *Il Sole 24Ore*, 5 June 2023. Available at: <https://www.ilsole24ore.com/art/nome-codice-giove-ecco-l-algoritmo-polizia-prevenire-reati-AEAoJIZD>; *Intelligenza artificiale: da Keycrime a Giove, nuovo investigatore per la Polizia*. Article published on *Il Sole 24Ore*, 10 June 2023. Available at: https://www.ilsole24ore.com/art/intelligenza-artificiale-keycrime-giove-nuovo-investigatore-la-polizia-AE9i1EbD?refresh_ce=1. K. CARBONI, *Cosa sappiamo di Giove, il software italiano di polizia predittiva*. Article published on *Wired*, 6 June 2023. Available at: <https://www.wired.it/article/giove-polizia-predittiva-software-italiano-come-funziona-garante-privacy/>.

¹⁸¹ See CIMMARUSTI, MAZZEI, *Nome in codice Giove*, cited.

¹⁸² Interrogazione Parlamentare, Atto n. 3-00499, 13 June 2023. Available at: <https://www.senato.it/japp/bgt/showdoc/frame.jsp?tipodoc=Sindisp&leg=19&id=1378767>. The requests to the Ministry are: «(a) what steps it intends to take to introduce the Giove system in Italy, whether there is other software of this kind already in use or the use of which is envisaged; (b) which companies have been involved in defining this technology, its implementation and its development; (c) what stage has been reached in talks with the data protection authority with regard to an assessment of the impact which the introduction of this system would entail (d) what level of detection would be possible and achievable without violating the privacy of individuals; (e) what would be the effects on urban planning in cities of such a penetrating and intrusive capacity to profile individuals and their behaviour, in the light of a very negative European and international debate on the use of such invasive technologies that infringe people's rights and pending a European decision that will make it compulsory to regulate such use, explicitly prohibiting the possibility of 'predictive policing'». Own translation from the Italian version.

In Italy, another predictive policing software in use is XLAW, developed within the Naples Police Department (Questura di Napoli)¹⁸³. The software is aimed at identifying criminal patterns focusing on a specific type of crime, namely street crimes, such as robberies and scams¹⁸⁴. To this end, different kind of data were collected – date, time, address, type of environment (private house, commercial building, etc.), type of author (age, gender, *modus operandi*), type of victim/s (age, gender, social and economic status, etc.), type of target (money, object, etc.) and also incentive factors (tourists, buses or trains arriving, etc.)¹⁸⁵ – and the heuristic method was adopted. In the end, the heuristic algorithm XLAW was created based on a supervised machine learning approach¹⁸⁶. Its purpose was to identify criminal patterns and strategies, identify the “hunting reserves”, when and where crimes were likely to occur, describe the type of author, *modus operandi*, type of target and type of victim¹⁸⁷. Finally, a risk-index of the city (PCrime) could also be generated¹⁸⁸. Compared to the UK, the Netherlands and Germany, Italy does not seem to invest much, especially in person-based predictive policing software. Already from this brief description, the differences between the various software are obvious: some use huge amounts of data, both personal and non-personal, others have access to rather limited data, usually police data; some rely on more complex models, others less so; some are more explainable, others less so.

2. Considerations on predictive policing

The aim of the preceding paragraphs was not to describe all existing predictive policing systems in Europe, but rather to describe their main features in order to introduce more general considerations on the issues faced in relation to the use of algorithmic tools for crime prediction and prevention¹⁸⁹.

Predictive policing involves using data-driven software to anticipate criminal activity, either by forecasting *where* crimes are likely to happen (place-based predictive policing) or *who* is more likely

¹⁸³ The development and functioning of XLAW is described in E. LOMBARDO, *Sicurezza 4P. Lo studio alla base del software XLAW per prevedere e prevenire i crimini*, Mazzanti Libri, 2019.

¹⁸⁴ Lombardo and his staff abandoned the concept of «criminal hot spots» and introduced the concept of «hunting reserve» (riserva di caccia), where desirable targets (people or places) and specific «hunting rules» can be identified. They found out that all street crimes present some common characteristics: i) they are concentrated in specific areas of the urban context, ii) the choice of the targeted area depends on opportunities and absence of adequate control, iii) criminals repeatedly act in the targeted area. The following step was to isolate criminal patterns and create a CrimeMap with all the hunting reserve. See LOMBARDO, *Sicurezza 4P*, cited, 75.

¹⁸⁵ *Ibidem*, 109.

¹⁸⁶ G. DI GENNARO, E. LOMBARDO, *Intelligenza artificiale e politiche di sicurezza urbana: verso quali modelli?*, in A.F. URICCHIO, G. RICCIO, U. RUFFOLO (Eds.), *Intelligenza Artificiale tra etica e diritti*, Cacucci Editore, 2020, 667.

¹⁸⁷ LOMBARDO, *Sicurezza 4P*, cited, 112.

¹⁸⁸ *Ibidem*, 115.

¹⁸⁹ For a general discussion on algorithms and ethics see B.D. MITTELSTADT, P. ALLO, M. TADDEO, S. WACHTER, L. FLORIDI, *The ethics of algorithms: Mapping the debate*, in *Big Data & Society*, 3(2), 2016, 1–21.

to be involved in a crime (person-based predictive policing). The development of such systems actually predates the recent wave of AI/ML popularity, and the underlying approaches vary widely. Before machine learning became mainstream, predictive policing software was largely based on statistical models and rule-based systems, such as hotspot mapping (e.g., CompStat in the 1990s), regression and statistical forecasting (drawing from theories like repeat victimization and routine activity theory) and heuristic-based rules (some early systems simply automated criminology insights). These systems relied on traditional criminological theories, GIS mapping, and basic statistical forecasting. With the rise and growing popularity of AI and data science in the late 2000s and 2010s, predictive policing tools began incorporating machine learning and more complex statistical modelling, such as pattern recognition. Instead of fixed statistical assumptions, ML models could “learn” correlations from large crime datasets, sometimes incorporating socioeconomic, demographic, and environmental variables. Today, predictive policing software is not uniformly based on AI: some systems are still built on traditional statistical forecasting, while others use modern ML pipelines, with differences in terms of complexity and adaptability.

Determining whether a predictive policing software qualifies as an AI system under the definition discussed in Chapter 1 is not a straightforward exercise. The most ambiguous category is place-based predictive policing. Such systems are certainly “machine-based”: they rely on software running on hardware, analyse data – such as crime reports, temporal and geographical information, and sometimes additional data – and generate forecasts on the likely time and location of future offences. Their outputs guide real-world decisions, such as patrol deployment, intensified surveillance, or stop-and-search practices, thus impacting on physical or virtual environments. Nonetheless, grey areas persist concerning the requirement of “inference”. On this point, the Commission’s *Guidelines on the definition of an artificial intelligence system* clarify that classical heuristic systems and simple prediction tools fall outside the scope of the AI Act. As far as classical heuristics are concerned, they are defined as “*problem-solving techniques that rely on experience-based methods to find approximate solutions efficiently*”¹⁹⁰. They typically use predefined rules, pattern-recognition shortcuts, or trial-and-error strategies, rather than adaptive, data-driven learning. As the Guidelines emphasise, while such methods may be effective, they lack the inference capacity characteristic of AI systems. Accordingly, a predictive policing tool that relies on traditional hot-spot analysis techniques – such as grid mapping, covering ellipses, kernel density estimation, or similar heuristic methods – may not, from my perspective, satisfy the criteria for classification as an AI system. For example, software XLAW is

¹⁹⁰ European Commission, *Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)*, cited, paragraph 48.

based on simple heuristics, thus it may fall outside the notion of AI systems under the AI Act. The same may hold true for systems based on the near-repeat approach, such as PredPol and PreMap. Likewise, basic risk-terrain modelling systems, which estimate crime risk primarily on the proximity to “risk-inducing” environmental features, may also fall outside the definition, though more recent and complex iterations of risk-terrain modelling – incorporating larger datasets and more sophisticated inference – may yield a different conclusion. By contrast, systems employing data-mining techniques to generate predictions are more likely to qualify as AI systems. Yet even here, machine-based systems that achieve their outputs through simple statistical learning rules may still fall outside the AI Act’s scope, despite technically relying on machine-learning. HunchLab 2.0, described in Chapter 2, with its greater complexity and enhanced predictive mechanisms, may satisfy the requirements to be considered an AI system. In any case, determining whether a given software system falls within the definition remains open to debate. The task is even more complex for software used by law enforcement authorities, where publicly available information is scarce, making it difficult to analyse the system’s specific characteristics and reach a well-grounded assessment.

Person-based predictive policing systems, by contrast, almost unequivocally fall within the AI Act’s definition of an AI system. They meet all the definitional criteria, and regulators explicitly treat them as high-risk AI applications due to their significant implications for fundamental rights. Systems such as the Strategic Subject Algorithm¹⁹¹ almost certainly fall within the definition of AI systems under the AI Act. In particular, it is a risk-scoring model, combining numerous variables (arrests, victimization, gang ties, network exposure), using weighted coefficients to compute a composite score, updating as new data entered the system and producing probabilistic estimations of involvement in gun violence. In a similar manner, hessenDATA may also be characterised as an AI system. In particular, hessenDATA is a data-integration and investigative analytics platform that, unlike traditional predictive-policing software, does not automatically generate risk scores or crime forecasts. Rather, it does infer new relationships from input data, which can subsequently be used by public authorities to guide operational decisions. For example, the system can automatically link persons, objects, phone numbers, and addresses across multiple databases, and perform network analyses that identify relationships between multiple entities. In any case, the current definition of an AI system still excludes certain types of software used for predictive policing. Although the systems falling outside the definition may not present the greatest risks to individual rights, they can nevertheless have

¹⁹¹ For a detailed analysis of the Strategic Subject List see A. TUCEK, *Constraining Big Brother: The Legal Deficiencies Surrounding Chicago’s Use of the Strategic Subject List*, published on *The University of Chicago Legal Forum*. Available at: <https://chicagounbound.uchicago.edu/uclf/vol2018/iss1/18/>.

negative effects, contributing to the over-policing of particular areas and, consequently, specific communities, with inevitable consequences in terms of feedback loops and potential discrimination.

2.1. Data and data protection concerns

Despite the differences in the techniques used to make predictions, the software described above raise concerns to a greater or lesser extent. One of the most critical aspects of predictive policing relates to data¹⁹². In some cases, like PRECOBS, XLAW or KeyCrime, data used to develop the software are crime-related data, collected and stored by the police¹⁹³. The types of data used are rather limited in their number and generally include geographical data, social, cultural and economic data. On the other hand, other predictive policing platforms, such as hessenDATA or POL-INTEL, use much more data for their development, including also data collected from private providers (e.g. social media platforms), and then acquired by the police¹⁹⁴.

The use of social media data – such as data on users (e.g. names, registration date, e-mail addresses) or users' interactions with the platform (e.g. posts and likes) – in the context of predictive policing may raise special concerns¹⁹⁵. Indeed, the fact that a person creates an account on a social media platform and publishes posts on it does not mean that the person has given his/her consent to use those data indiscriminately. Accordingly, an indiscriminate access to those data by the police should not be allowed, preferably requiring a court order that authorises access. The sensitive issue of police access to social media data was specifically addressed by the German Federal Constitutional Court in the February 2023 ruling, where the Court emphasised the need to exclude social network data from those

¹⁹² Data protection concerns are identified in almost all institutional documents. See Europol, *AI and policing*, cited; European Crime Prevention Network, *Artificial intelligence and predictive policing: risks and challenges*, June 2022. Available at: <https://eucpn.org/sites/default/files/document/files/PP%20%282%29.pdf>. UNICRI, Interpol, *Responsible AI Innovation in Law Enforcement. AI toolkit*, February 2024. Available at: [Artificial Intelligence Toolkit](#).

¹⁹³ EGBERT, LEESE, *Criminal Futures*, cited, 76.

¹⁹⁴ It is generally acknowledged that private companies generate profits not only through their primary business, but also from a «secondary business» which is connected to the selling of their customers' information to the so called «data brokers». Data brokering currently represents a highly profitable industry, as demonstrated by the literature on the topic. See R. KANWAL, K. WALBY, *Tracking the Surveillance and Information Practices of Data Brokers: A Report*, 4 July 2024. Available at: <https://www.uwinnipeg.ca/caij/docs/reports/tracking-the-surveillance-and-information-practices-of-data-brokers.pdf>. U. REVIGLIO, *The untamed and discreet role of data brokers in surveillance capitalism: a transnational and interdisciplinary overview*, in *Internet Policy Review*, 11(3), 2022, 1-27. Police departments might purchase data from data brokers. An example is the MET police that reported purchasing information from data brokers. Available at: <https://www.met.police.uk/foi-ai/metropolitan-police/disclosure-2024/january-2024/purchasing-information-data-brokers/>. Apart from data brokers selling individuals' data to police departments, there are legal obligations that require private companies to exchange data and information with law enforcement authorities.

¹⁹⁵ In France, for example, (some) social media data can be used by tax authorities in order to detect and punish tax frauds. See CNIL, *Délibération n° 2019-114 du 12 septembre 2019 portant avis sur le projet d'article 9 du projet de loi de finances pour 2020*.

accessed by hessenDATA in order to reduce interference with the right to informational self-determination¹⁹⁶.

When personal data are involved, data protection concerns may arise. Some types of software, such as those sold by Palantir, allow access to multiple databases in an interoperable manner, being precisely data integration platforms¹⁹⁷. Therefore, while the police could have access to all the data mentioned above even without these platforms, they would have access to them through different databases, which behave like silo structures. Accordingly, operators would be required to search each register or database and then merge the responses from the systems in order to obtain meaningful information. As stated by Brayne, Palantir platform – and other platforms like Palantir – “*helps overcome this fragmentation by integrating previously disparate data sources into a single search*”¹⁹⁸. Data protection issues arise from the fact that, through the integration of several databases, fundamental principles of data protection, especially the purpose limitation principle and the storage limitation principle, may be put in tension. Furthermore, it might be possible to re-identify the natural person to whom the data relate, even if the data have previously been rendered anonymous¹⁹⁹.

Where predictive policing software is used, a clear, specific, and sufficiently detailed legal basis for the processing of personal data is essential. It thus remains always necessary for the legislator (national or European) to introduce a legal basis that authorises the processing of personal data. A notable example in this regard is the judgment delivered on 16 February 2023 by the First Senate of the German Federal Constitutional Court (*Bundesverfassungsgericht*)²⁰⁰. In that decision, the Court examined two

¹⁹⁶ *Bundersverfassungsgericht*, February 16, 2023, cited, 44. It is noteworthy that in France, Article 154 of the Finance Act for 2020 (LOI n° 2019-1479 of 28 December 2019) initially authorised, on an experimental basis, the tax administration (DGFIP) and the customs and indirect taxes administration (DGDDI) to collect and exploit, through computerized and automated processing, content that is *manifestly made public* and publicly accessible on online platforms, including social networks, without the use of facial recognition, for the purpose of detecting certain serious tax and customs offences. Decree n° 2021-148 of 11 February 2021, adopted specifically to implement Article 154, laid down the practical modalities and safeguards governing such automated collection and exploitation of publicly accessible platform data. The Constitutional Council, in its Decision n° 2019-796 DC of 27 December 2019, reviewed Article 154, upheld its constitutionality subject to limits, and censured the provision allowing the system to be used for the less serious offence of late filing following a formal notice, striking out the words “*b et*” from the list of targeted breaches. Subsequently, the CNIL, in Délibération n° 2020-124 of 10 December 2020, issued an opinion on the implementing framework; this opinion, published in February 2021, was taken into account when finalizing the decree implementing Article 154. The scheme was later amended by Article 112 of the Finance Act for 2024 (LOI n° 2023-1322 of 29 December 2023), which led to the adoption of Decree n° 2024-1274 of 31 December 2024, updating the 2021 decree and revising the conditions and modalities for the collection and use of publicly accessible platform data.

¹⁹⁷ The kind of platform sold by Palantir was defined as a «*single sign-in solution for retrieving all [the] data*». GALIS, KARLSSON, *A world of Palantir*, cited, 2439.

¹⁹⁸ BRAYNE, *Predict and Surveil*, cited, 32.

¹⁹⁹ Moreover, absolute anonymity in the digital world is by no means easy to guarantee, as recognised by G. FINOCCHIARO, O. POLLICINO, *L'anonimato assoluto nel mondo digitale non si può garantire*. Article published on *Il Sole24Ore*, 8 November 2023.

²⁰⁰ *Bundersverfassungsgericht*, I Senate (President Stephan Harbarth), Judgement 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20. Available at: https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2023/02/rs20230216_1bvr154719en.html.

legislative provisions – respectively §25a of the Security and Public Order Act for the Land Hesse²⁰¹ and § 49 of the Act on Data Processing by the Police for the Land Hamburg²⁰² – which both empower the police to carry out automated data analysis or data interpretation²⁰³. More specifically, the relevant provisions authorise the police to employ automated data analysis (in Hesse) or automated data interpretation (in Hamburg) for the processing of stored personal data organised across multiple databases. As previously discussed, both provisions served as the legal basis for the Gotham operating system acquired by the German Länder from Palantir – subsequently renamed hessenDATA – a platform which, as anticipated, functions as an integrative system designed to enable interoperability across multiple databases²⁰⁴. In examining the implications of automated data analysis and interpretation tools, including AI-based systems, for the fundamental right to informational self-determination (*Recht auf informationelle Selbstbestimmung*)²⁰⁵, the Court held that such tools constitute a distinct and potentially more serious interference than the initial collection of personal data. Therefore, except where data are reused by the same authority, for the same task, and to protect the same legal interests that justified their original collection, automated processing requires separate constitutional justification and a specific statutory basis, in line with the principle of purpose limitation²⁰⁶. In particular, the Court noted that linking datasets and applying algorithmic techniques

²⁰¹ Hessisches Gesetz über die öffentliche Sicherheit und Ordnung – HSOG.

²⁰² Hamburgisches Gesetz über die Daten Verarbeitung der Polizei – HmbPolDVG.

²⁰³ More specifically, “*In view of the expanded technical possibilities for using information technology in police work, these provisions establish a specific legislative basis for linking previously unconnected, automated data bases and data sources within analysis platforms and for systematically mining the available datasets through searches, thereby enabling the police to carry out their tasks more efficiently and effectively*”. *Bundersverfassungsgericht*, cited, paragraph 3.

²⁰⁴ In the case of the Land Hesse, hessenDATA had automated access to three databases: POLAS (a police information system containing repressive law-enforcement data), ComVor (a case-management system used for all proceedings), and CRIME-ST (a case-management system designed to store preventive data for future investigations). Among the other sources to which hessenDATA had access was traffic data obtained through telecommunications surveillance and retrieval queries, data acquired through forensic extraction, such as information taken from mobile phones seized by the police, data from police telexes, a type of internal email system used by the police to exchange information across the Land of Hesse. Other sources to which hessenDATA had access, despite non automatically, was the source systems of other Länder, the Federation and foreign states, nor were any other public or non-public sources. As specified by the Ministry of the Interior, however, such data could be requested – within the applicable statutory limits – and the relevant sources could then be integrated into the platform for analysis. The same applied, in individual cases, to data obtained from police investigative measures carried out for preventive purposes, as well as to data arising from criminal investigation measures, such as telecommunications surveillance or the monitoring of private homes. See *Bundersverfassungsgericht*, Judgement, paragraph 11.

²⁰⁵ Art. 2(1) in conjunction with Art. 1(1) of the German Basic Law. Interference with the right depends, at least in part, by the severity of interference resulting from the preceding collection of the data; in this respect, the principles of purpose limitation and change in purpose apply. Moreover, automated data analysis or interpretation amount to a separate interference because the further processing of data can result in specific detrimental effects which might be more onerous than the severity of interference resulting from the original data collection; in this respect, the principle of proportionality in the strict sense requires that additional justification be provided.

²⁰⁶ Indeed, based on the principle of purpose limitation, authorising the automated analysis or interpretation of stored constitute a further use of previously collected data beyond the grounds that originally justified its collection. This constitutes a distinct interference with fundamental rights and therefore requires separate justification under constitutional law, in accordance with the principle of purpose limitation. Therefore, if the legislator permits a use of data beyond the specific grounds that prompted the original data collection and beyond the reasons that justified the original data collection,

can generate relationship profiles, infer new information not evident from isolated datasets, and produce investigative leads – even concerning individuals not previously under suspicion – thereby significantly intensifying the impact on fundamental rights. Moreover, the seriousness of the interference depends on factors such as the volume and type of data, the degree of interconnection, and the sophistication of the analytical methods. Legislative design is therefore crucial: the breadth of permissible purposes, the safeguards imposed, and the technical capacities authorised jointly determine whether the interference is constitutionally acceptable.

Importantly, the Court distinguished between data use that remains within the original purpose of collection and data use that involves a change of purpose. While the former may inform subsequent investigations, the latter triggers stricter constitutional requirements. The contested provisions permitted both forms of processing and allowed extensive analysis of large datasets without adequate distinctions regarding data sources or original purposes. Owing to the breadth of these powers and the heightened constitutional standards required, the Court held the provisions unconstitutional. The Court further stressed that the legislature must define the degree of interference by setting clear rules on the permissible types and scope of data, as well as limits on analytical methods. More serious interferences may be justified only under stringent conditions and for the protection of especially important legal interests, subject to a “sufficiently identifiable danger” threshold. Less severe interferences may rely on lower thresholds, depending on the weight of the legal interests at stake.

Data protection concerns are also one of the main reasons for the divestment of Omnia, the Norwegian predictive policing software sold and customised by Palantir. In particular, in 2016, the Norwegian police decided to purchase a software that would help them fulfil their obligations under the EU’s Prüm Agreement²⁰⁷, and Palantir was awarded the contract for €7.3 million through its Gotham platform. The software (then called “Omnia”) was supposed to connect data from nineteen different police registers²⁰⁸ and to facilitate interoperability between them enabling quick and easy searches across the databases through a single query. However, the Norwegian police announced the end of the project in 2020, before it became operational. The major tensions were between Palantir’s IT

it must create a separate statutory basis to that end. The contested provisions permit the further use of data not only for purposes in line with the original purpose but also for changed purposes. For a detailed discussion see *Bundersverfassungsgericht*, Judgement, paragraphs 57 and 58.

²⁰⁷ *Council Decision 2008/615/JHA of 23 June 2008 on the stepping up of cross-border cooperation, particularly in combating terrorism and cross-border crime*. Available at: <https://eur-lex.europa.eu/eli/dec/2008/615/oj/eng>. The Decision is aimed at improving cross-border cooperation between Member States’s police and judicial authorities. It focuses on the automated access to DNA profiles, dactyloscopy data and national vehicle registration data, the supply of data relating to major events, particularly those connected to terrorism and cross-border crime and the supply of information in order to prevent terrorist offences.

²⁰⁸ GUNDHUS, WATHNE, *Resistance to platformization*, cited. The name *Omnia* precisely recalls the concept of «everything», meaning that the software was supposed to connect multiple databases, from vehicles registers to DNA databases.

consultants and police Data Protection Officers (DPOs), considering their opposite visions on the project: the former were for the “use-and-adjust” approach, while the latter were for the “proof-of-concept” approach, thus requiring Palantir to satisfy all the requirements – such as in terms of data quality and transparency in the analysis – before the test and the use of the software²⁰⁹. Furthermore, there were differing views between Palantir’s IT consultants and DPOs on the consequences of poor data quality, the former being concerned only with the integration and interoperability aspects²¹⁰. Most importantly, there were different visions among the police DPOs on the data basis for the processing of personal data and the legitimacy of the data integration itself²¹¹. The Norwegian example is one that highlights the difficulty of using software that allows interoperability in a context – such as the European one – where data protection legislation is extremely rigorous. Another case in which data protection problems have been identified is the Gang Violence Matrix risk assessment software, developed by the Metropolitan Police to identify individuals and classify them according to the likelihood that they participate in gang activity. In 2018, after direct criticism of the programme, the Information Commissioner’s Office issued an enforcement notice, compelling the Metropolitan Police to bring it into compliance with data protection laws²¹².

Moreover, the fact that most predictive policing software is developed and then purchased by police authorities may imply the transfer or exchange of data collected and held by public bodies (*i.e.* the police) with private entities²¹³. The transferring of data from public-to-private actors deserves extreme caution, especially considering privacy legislation, and raises responsibility and accountability issues, which are evidently avoided with the development of the system in-house.

2.2. Bias and non-discrimination

Criticism that has been directed towards predictive policing is connected to the risk of these systems to perpetrate or even amplify biases that may lead to discrimination²¹⁴. The emergence of bias in

²⁰⁹ «This has to be thoroughly developed, thoroughly documented and thoroughly tested before you can start using some thing yourself, in line with the proof-of-concept approach». GUNDHUS, WATHNE, *Resistance to platformization*, cited, 9 reporting the wording of a Kripos employee.

²¹⁰ The idea that guided Palantir’s IT consultants was «the more data you had, the better the quality of analyses is».

²¹¹ GUNDHUS, WATHNE, *Resistance to platformization*, cited, 14.

²¹² ICO, *ICO finds Metropolitan Police Service’s Gangs Matrix breached data protection laws*, Information Commissioner’s office, 2018. Available at: <https://www.met.police.uk/SysSiteAssets/media/downloads/force-content/met/about-us/gangs-violence-matrix/ico-enforcement-notice.pdf>.

²¹³ MARCHETTI, *Voce Amministrazione Digitale*, cited, 100.

²¹⁴ On the specific topic of predictive policing and biases, see FRA, *Bias in algorithms. Artificial Intelligence and Discrimination*, Report, 2022; EUCPN, *Artificial intelligence and predictive policing: risks and challenges*, cited. G. GÜLTEKIN-VÁRKONYI, *Predictive Policing and Bias in a Nutshell: Technical and Practical Aspects of Personal Data Processing for Law Enforcement Purposes*, in K. KARSAI, A. SÖZÜER, L. WÖRNER (Eds.), *Digital Criminal Justice: A Studybook Selected Topics for Learners and Researchers*, 2022, 39-59. K. LUM, W. ISAAC, *To predict and serve?*, in *Significance Magazine*, 13(5), 2016, 14–19. W.D. HEAVEN, *Predictive policing algorithms are racist. They need to be*

predictive policing software is attributable to a multiplicity of factors, most notably the data drawn upon for the purposes of training the model or operationalising the system, and the criteria governing the selection and weighting of the predictors on which the system relies in producing its outputs.

Data used is generally historical data – such as data on past crimes, arrest or stop data or call for service data – related to specific populations within a particular timeframe. Therefore, the software risks being blind to context-specific political, social and economic aspects of the population in relation to which it will be used²¹⁵ and to specific political and social changes that eventually will happen in the future²¹⁶. Lack of context specificity may compromise the accuracy of the system in relation to different populations or a different time frame. Beyond population-specific characteristics, predictive software equally fails to account for individual specificities. Algorithmic risk scoring software, for instance, does not assess the individual as such, but instead draws on group-level data, proceeding on the assumption that individuals sharing a similar history will exhibit comparable behavioural patterns.

A further concern relates to the quality of police data, which cannot always be regarded as complete or reliable²¹⁷. This is due to under-reporting of crime or incorrect reporting of crime data, which eventually result in distorted data, false prediction and wrong decisions²¹⁸. As is well established in criminological scholarship, police data cannot be taken as a straightforward reflection of the reality of crime; they are, rather, a product of police activity and the multiple factors that condition it²¹⁹. Low reporting rates, for example, are influenced by multiple factors, such as personal characteristics, socio-

dismantled, MIT Technology Review, 2020. Available at: <https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/>. *Idem*, *Predictive policing is still racist—whatever data it uses*, MIT Technology Review, 2021. Available at: <https://www.technologyreview.com/2021/02/05/1017560/predictive-policing-racist-algorithmic-bias-data-crime-predpol/>.

²¹⁵ The importance of context-specific evaluation methodologies to ensure the scientific validity of statistical algorithms used by the police is stressed by BABUTA, OSWALD, *Data Analytics and Algorithms in Policing in England and Wales*, cited, 47.

²¹⁶ In their work MOSES, CHAN, *Algorithmic prediction in policing*, cited consider and criticise assumptions on which predictive policing is based, one of them is “*The future is like the past*”. This is an incorrect assumption, as the context is always changing, therefore the use of past data to make prediction may lead to inaccurate or incorrect results. In L. TORCHIA, *La giustizia amministrativa digitale*, in M. RAMAJOLI (Ed.), *Una giustizia amministrativa digitale?*, il Mulino, 2023, 52 in relation to predictive justice the author stress that the risk is not to predict the future but to project the past.

²¹⁷ A. BABUTA, M. OSWALD, C. RINIK, *Machine Learning Algorithms and Police Decision-Making. Legal, Ethical and Regulatory Challenge*, RUSI Whitehall Report 3-18, 2018, 8. C.J. MOSHER, T.D. MIETHE, T.C. HART, *The mismeasure of crime*, Thousand Oaks, Sage Publications, 2010. S.D. LEVITT, *The relationship between crime reporting and police: Implications for the use of Uniform Crime Reports*, in *Journal of Quantitative Criminology*, 14(1), 1998, 61–81. «[P]olice records do not measure crime. They measure some complex interaction between criminality, policing strategy, and community–police relations». LUM, ISAAC, *To predict and serve?*, cited, 16. To get an idea of the real crime rate, an interesting example is the victimization surveys, in which people are asked about their experiences of victimization and reporting to the police. In Italy, victimization surveys are conducted regularly by the National Institute of Statistics.

²¹⁸ Distorted data may also be the consequence of a late report of crime. In relation to burglaries, for example, people may not be awake or may not be at home while the crime happens. Under reporting of crime is not just a European concern. As far as the U.S. are concerned see Bureau of Justice Statistics, *Report: More Than Half of Violent Crimes Went Unreported to Police From 2006 – 2010*, RTI International, Press release, August 2012.

²¹⁹ JANSEN, *Data-driven policing: Negotiating the legitimacy of the police*, cited, 18.

economic background and trust in the police²²⁰. In addition, those who feel discriminated against by the police tend not to report crimes they witness or are victims of. Official crime statistics are further distorted by the differential observability of distinct offence types. Crimes with visible or tangible manifestations (such as burglary or robbery) are more amenable to detection and recording, while offences characterised by high levels of non-reporting, such as sexual violence, remain systematically underrepresented. The result is a structural bias in official data that reflects not only the incidence of crime, but the varying conditions under which different offences come to the attention of authorities. Eventually, the under or incorrect reporting of crime undermines the accuracy of the system²²¹.

Furthermore, data are not objective, meaning that they embed political, social and cultural values, priorities and practices, which can be biased²²². By way of illustration, individuals belonging to certain demographic groups may be subjected to disproportionately higher rates of stop and search, while police resources may be systematically directed towards the patrolling of specific neighbourhoods over others²²³. As a consequence, the data collected and recorded through such practices are inherently skewed towards those particular demographics and localities, which become structurally overrepresented in official databases²²⁴.

An extreme example of this consideration is related to potential illegal practices conducted by the police. In research conducted by Rashida Richardson, Jason M. Schultz and Kate Crawford²²⁵, the authors highlight questionable police data collection practices in several U.S. city police departments (among them, the NYPD and the Baltimore Police Department), highlighting systematic data manipulation, police corruption, falsification of police reports and other unconstitutional police practices²²⁶. They identified jurisdictions where there was an overlap in time between the development

²²⁰ See FRA, *Crime, safety and victims' rights – Fundamental Rights Survey*, Report, February 2021. FRA, *Second European Union Minorities and Discrimination Survey*, Publications Office of the European Union, 2018.

²²¹ MOSES, CHAN, *Algorithmic prediction in policing*, cited, 809. The authors describe the first assumption on which predictive policing is based: «the data used accurately reflect reality». As the authors recognise, this assumption is incorrect, as there is usually a gap between crimes committed and crimes reported, due to what individuals choose to report and what police authorities directly observe and record.

²²² This is widely recognised. See BRAYNE, *Predict and Surveil*, cited, 27. I.N. COFONE, *Algorithmic discrimination is an information problem*, in *Hastings Law Journal*, 70, 2019, 1389-1432. In LUM, ISAAC, *To predict and serve?*, cited, 14-19 the authors show how biased outcomes result from biased police-recorded data re-creating the functioning of PredPol.

²²³ FRA, *Your rights matter: Police stops - Fundamental Rights Survey*, Publications Office of the European Union, May 2021. In some jurisdictions, mainly but not solely in the U.S., places or neighbourhoods are closely intertwined with race. Consequently, the patrolling of certain areas means disproportionately impact on certain communities, such as those of colour. See also S. BASS, *Policing Space, Policing Race: Social Control Imperatives and Police Discretionary Decisions*, in *Social Justice*, 28(1), 2001, 156.

²²⁴ It is generally recognised that policing has a disproportionate impact on certain communities. See A.G. FERGUSON, *The Rise Of Big Data Policing: Surveillance, Race, And The Future Of Law Enforcement*, NYU Press, 2017, 131-136.

²²⁵ R. RICHARDSON, J.M. SCHULTZ, K. CRAWFORD, *Dirty data, bad predictions: how civil rights violations impact police data, predictive policing systems, and justice*, in *New York University Law Review*, 94, 2019, 192-233.

²²⁶ These practices were described in detail in J.A. ETERNO, E.B. SILVERMAN, *The crime numbers game: management by manipulation*, 2012 and later confirmed by an independent audit of the CompStat program. See D.N. KELLEY, S.L. MCCARTHY, *The Report of the Crime Reporting Review Committee to Commissioner Raymond W. Kelly Concerning*

and use of predictive policing systems and documentation on government investigations, consent decrees, memoranda of agreement or other evidence of corrupt, racially biased or otherwise unconstitutional police practices to see whether these “dirty data”²²⁷ were used to train or inform the software. As the authors point out, while the unconstitutional practices have been identified and stopped, reform of these processes and practices having been called for²²⁸, nothing or very little has been done with regard to the data that were collected through those practices. Where such data are fed into predictive policing systems without corrective intervention, the structural biases they embody are not neutralised but perpetuated, embedding discriminatory patterns into the algorithmic architecture itself and lending them an appearance of technical objectivity²²⁹.

Although the situation in Europe would appear to be different as to such unconstitutional and unethical practices are concerned, discriminatory policing practices have also been highlighted across the European territory²³⁰. As for now, it is sufficient to highlight that for better and unbiased software outcomes investing in updating and increasing the quality of datasets is one of the most important requirements²³¹.

Closely connected to the use of biased data to train a model are considerations on feedback loops. In general, a feedback loop occurs when predictions made by a system influence the reality in which the system is used, and this reality becomes the basis for the collection of other data that is then used to feed the same system²³². A biased outcome thus becomes the input data of the system, reinforcing and exacerbating biases over time. Feedback loops represent a critical aspect in predictive policing: a predictive policing software is used by police departments in order to allocate resources, namely which neighbourhoods should be patrolled by police officers; since more patrols are sent to the area identified

Compstat Auditing 41–43, 2013, http://www.nyc.gov/html/nypd/downloads/pdf/public_information/crime_reporting_review_committee_final_report_2013.pdf.

²²⁷ The authors refer to the concept of «dirty data» to identify not only «missing data, wrong data, and non-standard representations of the same data» but also «data that is derived from or influenced by corrupt, biased, and unlawful practices, including data that has been intentionally manipulated or “joked,” as well as data that is distorted by individual and societal biases». RICHARDSON, SCHULTZ, CRAWFORD, *Dirty data, bad predictions*, cited, 195.

²²⁸ As far as the NYPD is concerned unconstitutional practices were challenged in lawsuits against the NYPD which resulted in a federal court opinion. See *Floyd, et al. v. City of New York, et al.*, 959 F. Supp. 2d 540 (S.D.N.Y. 2013). As far as the Baltimore Police Department is concerned a consent decree was signed with the Department of Justice in order to stop the unconstitutional police practices. See *Consent Decree, United States v. Police Department of Baltimore City*, No. 1:17-cv-00099 JKB (D. Md. Jan. 12, 2017). Available at: <https://www.justice.gov/opa/file/925056/download>.

²²⁹ As the authors point out further evaluations of the extend of bias were prevented by the Government and corporates’ lack of transparency. RICHARDSON, SCHULTZ, CRAWFORD, *Dirty data, bad predictions*, cited, 198.

²³⁰ P. WILLIAMS, E. KIND, *Data-Driven Policing: The Hardwiring Of Discriminatory Policing Practices Across Europe*, Report, European Network Against Racism, 2019.

²³¹ For a discussion of the principle «bias in, bias out» in relation to the police and criminal justice sector see S.G. MAYSON, *Bias In, Bias Out*, in *The Yale Law Journal*, 128(8), 2019, 2218–2298. For more general discussion, see B. HANSON, S. STALL, J. CUTCHER GERSHENFELD, K. VROUWENVELDER, C. WIRZ, Y. RAO, G. PENG, *Garbage in, garbage out: mitigating risks and maximizing benefits of AI in research*, in *Nature*, 623(2), 2023, 28–31.

²³² FRA, *Bias in algorithms*, cited, 29.

as high risk, it is plausible that more crimes are actually recorded than in an area classified as low risk where no police are deployed; the record is then fed back into the system, thus reinforcing the high risk score identified by the system for this area²³³. The risk, therefore, is that predictive policing systems operate so as to reinforce and intensify the over-policing of certain neighbourhoods — typically those inhabited by disadvantaged and marginalised communities, which are already subject to disproportionate police scrutiny, while simultaneously contributing to the under-policing of other areas²³⁴. Crucially, this outcome bears little or no relation to the actual distribution of criminal activity. It is generally acknowledged that both over policing – which determines police stops and searches or identity checks – and under policing have adverse effects on fundamental human rights²³⁵.

Although person-based predictive policing directly affects fundamental rights and freedoms, place-based predictive policing can also have significant – albeit sometimes indirect – impacts on individuals. When a system is used to predict locations where crimes are likely to occur, any bias in the data can translate into biased outcomes. For example, if a neighbourhood is repeatedly flagged as high risk, residents there may experience heightened police presence, leading to more frequent stops, interventions, and surveillance, even if they themselves are not engaged in criminal activity.

Biased outcomes may depend also on the selection and weighting of indicators for predicting the risk score. All predictive policing systems focus on a limited, even if sometimes large, set of variables. For both person-based and place-based predictive policing systems, the choice of predictors (*i.e.* indicators that have predictive power) to be considered must be made carefully. In fact, the addition or omission of variables may affect both the accuracy and the fairness of the predictive software²³⁶.

The Dutch Crime Anticipation System (CAS) represents an illustrative example. Among the variables incorporated into CAS, several were of a particularly sensitive nature, including indicators such as the “Number of non-Western allochthones in zip code area” or the “Number of social benefits recipients

²³³ Helga Nowotny talks about the so-called «self-fulfilling prophecy». H. NOWOTNY, *In AI We Trust: Power, Illusion and Control of Predictive Algorithms*, Polity Press, 2021.

²³⁴ An interesting simulation of feedback loops in policing can be found in FRA, *Bias in algorithms*, cited, 38-45. Conclusions that are drawn from the simulation are the following: «[1] *The initial distribution of patrols seems to be the least relevant factor for the formation of feedback loops compared with the other relevant parameters.* [2] *The internal bias of machine learning models accelerates the formation of feedback loops; even in the situations in which no real differences could be amplified, it captures random variation, creates fictitious differences and progressively amplifies them.* [3] *When crime reporting rates are equal across neighbourhoods, they play an important role in mitigating the formation of feedback loops; the extent of the mitigation depends on the true crime rates. However, erroneous allocation of police patrols may occur when crime reporting rates differ significantly across neighbourhoods but the true distribution of crime is close to uniform.* [4] *Erroneous allocation of police patrols may also occur when crime observability differs across neighbourhoods.*».

²³⁵ FRA, *Bias in algorithms*, cited, 31. While over policing may affect fundamental rights because people are stopped, searched or even arrested, under policing may affect fundamental rights because the police fail to protect individuals from criminal acts of another individual. The case law of the ECtHR is huge. The leading case is ECtHR, Judgement of 28 October 1998, *Osman v. the United Kingdom*, Merits, No. 23452/94.

²³⁶ MOSES, CHAN, *Algorithmic prediction in policing*, cited, 811.

in zip code area”, proxies that explicitly encode ethnicity and socioeconomic status as predictive factors. In relation to the former, the term *allochthones*, as opposite to *autochthone*, is used to identify people with at least one foreign-born parent (by contrast, *autochthone* describes a person with both parents born in the Netherlands) and the reference to “non-Western” refers to South American, African and Asian countries, except Japan and Indonesia²³⁷. This indicator is thus directly related to the racial-ethnic origins of individuals living in the area. Therefore, it is not surprising that in 2017 the Dutch police removed this indicator from CAS. Similarly, the indicator that focuses on the number of benefit recipients in the area can be considered sensitive in nature, as it appears that people receiving benefits (*i.e.*, people with low incomes) are more likely to exhibit criminal behaviour²³⁸.

The CAS example shows that the choice of predictors is far from simple, considering that using some indicators, although they may be positively correlated with criminal behaviour, are considered unethical because they may lead to unfair decisions²³⁹. To avoid discriminatory outcomes, the omission of variables that describe sensitive information (such as race or ethnicity) may be justified. However, as evidenced by CAS, omitting such variables may not be sufficient, as other variables may be correlated with sensitive characteristics and act as proxies for them. This is the case with postcodes, which, especially in the U.S. context, identify an area where specific communities live, thus providing a proxy for race. Nor is it possible to avoid all variables related to sensitive characteristics which are statistically relevant, otherwise the accuracy of the system would be greatly reduced²⁴⁰. What should be acknowledged is that predictors used to build the model – rather than the model itself – are the central concern for feedback loops and perpetration of bias²⁴¹.

²³⁷ See OOSTERLOO, VAN SCHIE, *The Politics and Biases of the “Crime Anticipation System” of the Dutch Police*, cited, 6. It is also unclear why Indonesia and Japan are excluded from the «non-Western» countries.

²³⁸ Similarly, in the Netherlands, an automated welfare fraud detection system called SyRI was used leading to a globally known scandal. Hague District Court, NJCM et al. v The Dutch State, 5 February 2020. Available at: <https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBDHA:2020:1878>. For an interesting discussion on the judgement see M. VAN BEKKUM, F.Z. BORGESIU, *Digital welfare fraud detection and the Dutch SyRI judgment*, in *European Journal of Social Security*, 2021, 23(4), 323–340.

²³⁹ The COMPAS case is a relevant example. The ProPublica investigation suggested that the selection and weighting of variables contributed to the biased outcomes of the software. J. LARSON, S. MATTU, L. KIRCHNER, J. ANGWIN, *How We Analyzed the COMPAS Recidivism Algorithm*, May 2016. Available at: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>.

²⁴⁰ Among others, see T. CALDERS, S. VERWER, *Three naive Bayes approaches for discrimination-free classification*. *Data mining and knowledge discovery*, 21(2), 2010, 277–292; T. CALDERS, I. ŽLIOBAITĖ, *Why unbiased computational processes can lead to discriminative decision procedures*, in B. CUSTERS, T. CALDERS, B. SCHERMER, T. ZARSKY (Eds.), *Discrimination and privacy in the information society: data mining and profiling in large databases*, Springer, 2013, 43–60. F. KAMIRAN, T. CALDERS, M. PECHENIZKIY, *Techniques for discrimination-free predictive models*, in B. CUSTERS, T. CALDERS, B. SCHERMER, T. ZARSKY (Eds.), *Discrimination and privacy in the information society: data mining and profiling in large databases*, Springer, 2013, 223–240.

²⁴¹ OSWALD, GRACE, URWIN, BARNES, *Algorithmic risk assessment policing models*, cited, 229.

2.3.Errors and accuracy concerns

Another aspect to be considered is that predictive policing software almost inevitably produces errors. There are no prediction models that achieve one hundred per cent accuracy. Sartor identifies several reasons for this²⁴²: first, there are no methods that allow the programmer to identify every error in every possible piece of software; second, it is not possible to assess the correctness of the algorithm on all possible input data; third, an algorithm can be correct in most cases and still produce errors in relation to specific inputs. Absolute accuracy of prediction models is only possible when applied to the data used to build them. However, prediction models are useful precisely when applied to new and different data, so a percentage of inaccuracy must generally be accepted. Complex ethical and legal issues then arise, starting with the percentage of accuracy considered acceptable and the predictive model considered fair.

When the degree of accuracy is not one hundred per cent, errors occur. Such errors are referred to as “false positives” and “false negatives”²⁴³. The choice between which one is better is not objective, rather it is based on a “value-judgement”²⁴⁴. In the case of predictive policing software, for example, a police department may consider that predicting a person as “low risk” but actually “high risk” is not a desirable error, therefore overestimation of the person’s risk level is considered a better outcome, as it maximizes public safety. Therefore, false positives are preferred. By contrast, other police departments may consider the overestimation of a person’s risk level as a dangerous error, thus favouring underestimation of the person’s risk level. A person is thus predicted as “low risk” but actually “high risk” and false negatives are preferred. The most desirable errors have been referred to as “cautious errors”, while the least desirable ones as “dangerous errors”²⁴⁵.

Since completely avoiding errors is impossible, it is necessary to identify approaches that correctly distribute the type of prediction errors that occur. This is what the Durham Constabulary tried to do with the HART software using random forests²⁴⁶. In the case of HART, the prediction model overestimates the individual’s risk level. This means that proportionally more high-risk predictions are intentionally inaccurate and do not represent the actual risk level. Ethical questions may legitimately

²⁴² SARTOR, *L’informatica giuridica e le tecnologie dell’informazione*, 99.

²⁴³ LAGIOIA, ROVATTI, SARTOR, *Algorithmic fairness through group parities?*, cited, 465. False positives are described as «the individuals for whom the prediction is positive (reoffence is predicted) but that prediction is false (the individual will not reoffend)» while false negatives are described as the «individuals for which the prediction is negative (no reoffence is predicted) but that prediction is false (the individual will reoffend)».

²⁴⁴ F. KRAEMER, K. VAN OVERVELD, M. PETERSON, *Is there an ethics of algorithms?*, in *Ethics and Information Technology*, 13, 2011, 251–260.

²⁴⁵ OSWALD, GRACE, URWIN, BARNES, *Algorithmic risk assessment policing models*, cited, 229, 230.

²⁴⁶ OSWALD, GRACE, URWIN, BARNES, *Algorithmic risk assessment policing models*, cited, 229.

arise²⁴⁷. However, the severity of the consequences of errors must be taken into account. In some cases, such as in relation to HART software, the consequences are not so severe: if an individual is considered high risk, he or she will simply lose the opportunity to be considered eligible for the Checkpoint programme. In other cases, the consequences of this type of error can be extremely serious and lead to detentions and searches, arrests and deprivation of liberty.

The issue of overestimation and underestimation can be framed as the identification of a “trigger for action”²⁴⁸, that is the determination of a threshold at which a certain predicted probability justifies a specific classification or intervention. For instance, labeling individuals with a given risk percentage of reoffending as “high risk” rather than “moderate risk” entails a classification task, which involves selecting a cutoff value (e.g., 80%) to serve as this action threshold. This decision is inherently normative, as it reflects underlying objectives, such as minimizing costs or maximizing public safety, and requires explicit alignment of the system with those goals. In contexts such as policing and criminal justice, where the consequences of classification are substantial, the choice of threshold must be made with particular care. All instances falling below the threshold are classified as “negatives”, while those above are labeled “positives”; within each category, however, the model inevitably produces a mix of true positives/negatives and false positives/negatives, thus leading to accuracy issues.

2.4. Transparency issues

As previously observed, predictive policing software is not all the same, they are based on different types of data (e.g. historic crime data, environmental data, socio-economic data) and different amount of data, different methodologies (e.g. hot spot analysis, risk terrain modelling, data mining techniques) and different algorithms. Even though current predictive policing systems are not extremely complex, systems that depend on increasingly complex machine learning algorithms will soon be developed in Europe²⁴⁹, with negative consequences in terms of transparency and explainability of their outcomes. Predictive policing software which is based on complex algorithms, especially when machine learning and deep learning are concerned, suffer from opacity and are therefore described as “black-box”²⁵⁰, as it is not possible to provide a comprehensive explanation of the relationship between inputs and

²⁴⁷ *Ibidem*, 230. However, the rise of ethical questions depends on the public perception as to which choice is considered more ethical, that of increasing the protection of the community or that of being more cautious with the individual.

²⁴⁸ LAGIOIA, ROVATTI, SARTOR, *Algorithmic fairness through group parities?*, cited. The authors define the identification of the «trigger for action» as the second phase, the «classification phase», of the decision-making process.

²⁴⁹ EGBERT, LEESE, *Criminal Futures*, cited, 3. This is already the case with the software sold by Palantir.

²⁵⁰ The term can be traced back to F. PASQUALE, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, 2015.

outputs. The “black box problem” is extremely challenging when AI systems are used in the public sector²⁵¹, and the use of opaque AI systems in the specific area of law enforcement is even more critical, as there is no formal procedure preceding the decision and an obligation to state reasons is not even provided for by law²⁵². Moreover, the opacity of the model is aggravated by the general tendency to trust the software, without having the ability to question its results²⁵³.

Transparency problems may also arise in relation to those developing the predictive policing software. Indeed, some predictive policing systems are developed internally by police departments – such as PRECOBS, KeyCrime and XLAW – usually with the help of universities or research centres and then distributed locally or nationally. However, in most cases, predictive policing software is developed by private companies and subsequently purchased by police departments, as happened in Germany, Norway and Denmark, respectively with the acquisition of hessenDATA, Omnia and POL-INTEL from Palantir²⁵⁴. The distinction between contracting out or in-house development of the software is thus extremely relevant, as with the outsourcing of the software concerns related to proprietary software and trade secrets may arise²⁵⁵. In fact, very few companies publicly release their algorithms²⁵⁶ or openly describe which algorithms they use, what is the underlying functioning of the predictive model, what data it is trained on. The lack of transparency by private vendors over the systems they develop is extremely critical, since the public authority lacks the necessary information to use the software correctly²⁵⁷. Andrew Guthrie Ferguson described this situation as a “technological barrier”²⁵⁸ that prevents police, and public administrations in general, from the full comprehension of the systems they use with the consequence of blindly following the outcomes of the software²⁵⁹.

Another aspect that must be considered when evaluating transparency of predictive policing systems concerns police tactics²⁶⁰. For a system to be effective, the police do not usually reveal too much

²⁵¹ This is true even though the law enforcement sector is traditionally more opaque than other public sectors.

²⁵² See ARMIENTO, *Pubbliche amministrazioni e intelligenza artificiale*, cited, 190.

²⁵³ A.G. FERGUSON, *Illuminating Black Data Policing*, in *Ohio State Journal Of Criminal Law*, 15, 2018, 510.

²⁵⁴ «What is new and important about the current big data age and organisational form of policing we find ourselves in is the role of private actors in public policing». BRAYNE, *Predict and Surveil*, cited, 18.

²⁵⁵ MARCHETTI, *Voce Amministrazione Digitale*, cited, 100. For a discussion on the difference between proprietary software and open-source software see SARTOR, *L'informatica giuridica e le tecnologie dell'informazione*, cited, 119. See also S. DEL GATTO, *I sistemi proprietari, l'open source e la pubblica amministrazione*, in *Giornale di diritto amministrativo*, 5, 2021, 571-577.

²⁵⁶ PredPol is one of the few companies to have publicly described its algorithm. Predpol Inc., *Best Practices and Training Guide*, cited.

²⁵⁷ Some authors approached the topic from an even different perspective, questioning whether the implementation of the algorithm by a private party could result in an outsourcing of the administrative function itself. See A. DI MARTINO, *L'amministrazione per algoritmi ed i pericoli del cambiamento in atto*, in *Il diritto dell'economia*, 103(3), 2020, 628-629.

²⁵⁸ FERGUSON, *Illuminating Black Data Policing*, cited, 511.

²⁵⁹ These concerns, however, emerge whenever public administrations are called upon to decide which is the best alternative for them between outsourcing or in-house technology solutions.

²⁶⁰ FERGUSON, *Illuminating Black Data Policing*, cited, 513.

information about the systems they use and how they function²⁶¹. In fact, the release of detailed information on the functioning of the system could undermine the very rationale for its introduction. However, current obscurity in the use of such software by the police certainly has a negative impact on public trust and legitimacy of the police's authority²⁶². This is precisely what happened with the Italian software Giove. The public became aware of the development of this software thanks to media reports²⁶³ and subsequently a group of parliamentarians formally requested information on the development of the software from the Public Security Department of the Ministry of the Interior. However, no formal reply was provided. Lack of transparency also has negative consequences in relation to the principle of accountability. The police, as other public authorities, must remain accountable for their actions and decisions and the use of predictive policing software shall not alter this duty. Lack of transparency and explainability raises serious concerns in terms of the decision-maker's ability to justify its decision and the individual's ability to challenge the decision. Therefore, a combination of different approaches, both from a technical and legal point of view, is required to shed light on opaque systems.

²⁶¹ Police practices are traditionally closed from publicity. For a discussion on the lack of transparency in police practices see R. HARMON, *Why Do We (Still) Lack Data on Policing?*, in *Marquette Law Review*, 96(4), 2013, 1119-1146.

²⁶² The strict relationship between transparency and trust was highlighted in European Commission, *White Paper on Artificial Intelligence: a European approach to excellence and trust*, Brussels, 19 February 2020, COM(2020) 65 final. Available at: https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

²⁶³ *Intelligenza artificiale: da Keycrime a Giove, nuovo investigatore per la Polizia*. Article published on *Il Sole24Ore*, 10 June 2023. Available at: https://www.ilsole24ore.com/art/intelligenza-artificiale-keycrime-giove-nuovo-investigatore-la-polizia-AE9i1EbD?refresh_ce=1.

Part II: Automated video surveillance and face recognition technology

1. The police use of video analytics

In recent years, there has been an increasing interest of police departments in video surveillance systems enhanced by AI. From a general point of view, video analytics refers to the automated analysis of video footage to extract, interpret and understand data and get meaningful insights that can trigger actions or alerts, both in real-time and in delayed mode. Modern video analytics generally uses computer vision, a sub-field of AI, to obtain better results²⁶⁴. Some of the most common applications of AI for video analytics is object detection²⁶⁵, individual's identification, categorization and tracking, movement analysis²⁶⁶, emotion recognition and anomaly detection²⁶⁷. In all these scenarios, developing a fully deterministic approach proves challenging due to the complexity and variability of real-world conditions. Consequently, the use of machine learning or deep learning becomes essential, as they are better suited to handle large volumes of data and adapt to fuzzy patterns²⁶⁸.

The terms “algorithmic video surveillance”²⁶⁹, “intelligent” or “augmented video surveillance”²⁷⁰ are used to identify those applications which involve the integration of computer vision algorithms (or video image processing techniques²⁷¹) into video surveillance systems – existing devices (e.g. CCTV cameras) or *ad hoc* devices – to automatically detect and recognise objects, individuals, movements, events, or even emotions²⁷². In the following sections I analyse different types of technologies that use

²⁶⁴ Computer vision is based on deep learning and specifically on Deep Convolutional Neural Networks (DCNNs). Recently, Vision Transformers (ViT) has emerged as a competitive alternative to DCNNs.

²⁶⁵ For a technical analysis of object detection see M. MONIKA, U. RAJENDER, A. TAMIZHSELVI, A.S RUMALE, *Real-Time Object Detection In Videos Using Deep Learning Models*, in *ICTACT Journal on Image and Video Processing*, 14(2), 2023, 3103-3109.

²⁶⁶ For a technical discussion on movement analysis see H.-B. ZHANG, Y.-X. ZHANG, B. ZHONG, Q. LEI, L. YANG, J.-X. DU, D.-S. CHEN, *A Comprehensive Survey of Vision-Based Human Action Recognition Methods*, in *Sensors*, 19(5), 2019, 1005.

²⁶⁷ For a technical discussion on anomaly detection see R. CHALAPATHY, S. CHAWLA, *Deep Learning For Anomaly Detection: A Survey*, in *arXiv:1901.03407v2*, 24 January 2019.

²⁶⁸ S. AZRIA, F. WICKERT, *Facial Recognition: Current Situation and Challenges*, Consultative Committee Convention 108, November 13, 2019, 2.

²⁶⁹ La Quadrature du Net, *Algorithmic Video Surveillance. Dangers And Counter-Attacks*. Available at: <https://www.laquadrature.net/en/avs/>.

²⁷⁰ This term was used by the CNIL in its position on the use of image processing-enhanced devices in public places, setting out its reflections and analysis on the subject from an ethical, technical and legal perspective. CNIL, *Caméras Dites «Intelligentes» Ou «Augmentées» Dans Les Espaces Publics*, November 2024. Available at: <https://www.cnil.fr/fr/cameras-augmentees-espaces-publics>.

²⁷¹ See CNIL, *Facial Recognition: For A Debate Living Up To The Challenges*, 4 November 2019. Available at: <https://www.cnil.fr/en/facial-recognition-debate-living-challenges>.

²⁷² Adrien Tallent, a Ph.D. candidate in political philosophy, explains that «The argument in favor of AVS is that it could automatically detect behaviors, suspicious objects, or crowd movements without resorting to biometric surveillance, which relies on sensitive data like facial recognition, currently prohibited by law [...] However, according to some critics, AVS could be a disguised form of biometric surveillance since a person's gait or clothing might indirectly identify them» January

images as the focus of their processing, highlighting some differences that are necessary for a proper legal discourse, that I will address in detail in the following Chapters.

1.1. Image processing through Automated Video Surveillance

A necessary initial clarification is that not all intelligent video surveillance systems incorporate biometric techniques²⁷³. For this Thesis, I use the broader term “automated video surveillance” (hereinafter “AVS”) to distinguish these systems from those that are specifically designed for the processing of biometric data. AVS can serve multiple functions: to flag changes in the size of a crowd or in people’s movements, to detect the presence of objects (e.g. an abandoned suitcase in a train station) or specific events (e.g. a fight in a crowd), to identify the presence or the use of weapons (e.g. a person carrying a gun)²⁷⁴. Despite the absence of biometric technologies aimed at uniquely identifying the individual, AVS systems pose quite serious concerns compared to conventional ones, since the embedding of computer vision algorithms in surveillance devices expands the information that can be retrieved from a video stream²⁷⁵. More precisely, AI-enhanced surveillance cameras enable the police to automate the monitoring of video feeds and automatically retrieve data from them that respond to specific queries (e.g. searching for a specific event or object)²⁷⁶. This reduces the problem

2025. Available at: <https://www.sorbonne-universite.fr/en/dossiers/intelligence-artificielle/algorithmic-video-surveillance-will-enhanced-security-compromise-freedom>.

²⁷³ This distinction was stressed also by CNIL, *Facial Recognition*, cited, 4. An example is represented by tools developed during the Covid-19 pandemic which were used to detect mask in public spaces. The CNIL said they were not FRT-based tools since they were not technically capable of identifying individuals. CNIL, Délibération n° 2020-136 du 17 décembre 2020 portant avis sur un projet de décret relatif au recours à la vidéo intelligente pour mesurer le taux de port de masque dans les transports, para. 8. In relation to the management of the Covid-19 health crisis the CNIL issued some recommendations in relation to specific devices spreading in that period. CNIL, *La CNIL appelle à la vigilance sur l'utilisation des caméras dites «intelligentes» et des caméras thermiques*, June 2020. Available at: <https://www.cnil.fr/fr/la-cnil-appelle-la-vigilance-sur-lutilisation-des-cameras-dites-intelligentes-et-des-cameras>.

²⁷⁴ CNIL, *Caméras Dites «Intelligentes» Ou «Augmentées» Dans Les Espaces Publics*, cited, 6. Among the multiple use cases, those which are relevant for the purpose of this Thesis are: «Automated detection of situations that give rise to a presumption of infringement [...] or suspicious or potentially dangerous events (gathering of individuals, abnormally long presence of a person in a given place and at a given time, facial expression, behaviour expressing a state of anxiety)». Own translation.

²⁷⁵ As the CNIL recognises, «Such algorithms multiply the capabilities of conventional video devices». CNIL, *Caméras Dites «Intelligentes» Ou «Augmentées» Dans Les Espaces Publics*, cited, 5. Own translation. «Automate in order to deduce, probabilistically, certain information that may, where appropriate, enable decisions or concrete measures to be taken concerning them. Such a change does not constitute a simple technological evolution of video devices, but a change in their nature». *Ibidem*, 9. Own translation.

²⁷⁶ At the same time, surveillance cameras equipped with AI give police authorities new powers that allow for more thorough control over places and people, with negative consequences for individual rights and freedoms. La Quadrature du Net, *Algorithmic Video Surveillance*, cited. M. FOUCAULT, *Sécurité, territoire, population: Cours au Collège de France 1977-1978*, Hautes Etudes, 2004, 62-64. M. FOUCAULT, *Surveiller et punir, Naissance de la prison*, Gallimard, 1975, 215.

of human and material limitations that make the monitoring of all video streams unlikely, given the now large number of video surveillance cameras distributed throughout the city²⁷⁷.

As anticipated, information derived from intelligent video analysis may form the basis of subsequent decisions or actions taken by human operators. Inevitably, the use of AVS systems as the basis for decisions on an individual level (e.g. an augmented video surveillance system to detect crimes and help catch the perpetrator) is more intrusive than its use to obtain statistical information (e.g. to estimate the number of people using a ticket to access the metro). Therefore, the impact of such systems on individuals' rights and freedoms depends on multiple variables: the purpose pursued by the AVS system, the type of subsequent decisions or actions taken by human operators based on the system's outputs and the nature of data processed²⁷⁸.

Examples of AVS have rapidly spread throughout Europe. One of the most interesting cases in the European context is the one involving the 2024 Olympic and Paralympic Games (hereinafter "2024 Olympics") hosted in Paris. In 2022, the French government adopted a draft law that, among other provisions, established the legal framework for the experimental implementation of automated analysis of images from video surveillance cameras. Images analysed are those collected in relation to sports, recreational and cultural events (e.g. music festivals, concerts, football matches) which are particularly exposed to the risk of acts of terrorism or serious threats to personal security, due to their large scale ("large-scale events"). In December 2022, the CNIL issued an opinion on the proposed legal framework²⁷⁹, recalling its position on augmented cameras published in the same year²⁸⁰. Although the CNIL emphasised the importance of experimentation in relation to the use of algorithmic processing, it placed emphasis on the conditions under which such experimentation must take place²⁸¹. At the same time, after the CNIL had taken a position, the Conseil d'État issued an opinion on the draft law²⁸², leading the French Government to amend the draft law.

²⁷⁷ CNIL, *Caméras Dites «Intelligentes» Ou «Augmentées» Dans Les Espaces Publics*, cited, 9. As recognised by the CNIL, the increase in datasets currently makes manual analysis almost impossible. Similarly, La Quadrature du Net, *Algorithmic Video Surveillance*, cited, 18-19.

²⁷⁸ CNIL, CNIL, *La CNIL appelle à la vigilance sur l'utilisation des caméras dites «intelligentes» et des caméras thermiques*, cited, 11.

²⁷⁹ CNIL, Délibération n° 2022-118 du 8 décembre 2022 portant avis sur un projet de loi portant sur les jeux Olympiques et Paralympiques de 2024 (demande d'avis n° 22017438).

²⁸⁰ CNIL, *Caméras Dites «Intelligentes» Ou «Augmentées» Dans Les Espaces Publics*, cited.

²⁸¹ Precisely, the CNIL identified the following conditions: compliance with data protection legislation, evaluation report, data protection impact assessment, specific purpose of the processing operations.

²⁸² Conseil d'État, Avis sur un Projet de Loi relatif aux jeux Olympiques et Paralympiques de 2024 et portant diverses autres dispositions, December 15, 2022. Available at: https://www.legifrance.gouv.fr/contenu/Media/files/autour-de-la-loi/legislatif-et-reglementaire/les-avis-du-conseil-d-etat-rendus-sur-les-projets-de-loi/2022/avis_ce_spox2233026l_cm_22.12.2022.pdf.

In May 2023, the French legislature finally adopted Law No. 2023-380²⁸³. Article 10 of the Law authorised the French Ministry of Interior during the 2024 Olympics²⁸⁴, as well as in other large-scale events, to use “augmented cameras”, defined by CNIL as “*video devices with associated software for automatic image analysis (algorithmic processing)*”²⁸⁵ which “*can detect or ‘recognize’ specific objects, behaviors or events*”²⁸⁶. It is noteworthy that these augmented cameras do not use face recognition or other biometric techniques²⁸⁷, rather they are solely supposed to automatically detect specific types of situations²⁸⁸, namely: i) the presence of abandoned objects; ii) the presence or use of weapons; iii) the failure of a person or vehicle to follow the common direction of traffic; iv) the crossing or presence of a person or vehicle in a prohibited or sensitive area; v) the presence of a person on the ground following a fall; vi) a crowd movement; vii) too high density of people; viii) a start of fire. Once one of these events has been detected, the system sends an alert to authorised agents who check the alert and validate it or not, deciding what actions to take next.

An evaluation committee has been created to evaluate the trials²⁸⁹. A first real-life test of augmented cameras was organized by the Paris Police during the Depeche Mode concert at the Accor Arena in Paris in March 2024. As reported by the CNIL, “*the device was able to check crowd density, detect people or vehicles crossing into prohibited areas, fire starts*”²⁹⁰. After positive tests, those technical solutions were effectively used during the 2024 Olympics²⁹¹. The decision to use AI-based surveillance technologies to ensure security and public order has been severely criticised by civil society and NGOs, who argued that this experimental and temporary use of AVS tested in Paris would become a

²⁸³ LOI n° 2023-380 du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024 et portant diverses autres dispositions (Law No. 2023-380), Article 10. Available at: <https://www.legifrance.gouv.fr/dossierlegislatif/JORFDOLE000046777392/>.

²⁸⁴ For an analysis on security arrangements during the Olympics from 1970s onwards see J. COAFFEE, *Evolving security motifs, Olympic spectacle and urban planning legacy: from militarization to security-by-design*, in *Planning Perspectives*, 39(3), 2024, 637-657. The topic is touched also by P. BOYLE, *Securing the Olympic Games: Exemplifications of Global Governance*, in H.J. LENSKEY, S. WAGG (Eds), *The Palgrave Handbook of Olympic Studies*, Palgrave Macmillan, 2022, 394-409.

²⁸⁵ CNIL, *2024 Olympics: CNIL's Q&A on your privacy and freedoms*. Available at: <https://www.cnil.fr/en/2024-olympics-cnils-qa-your-privacy-and-freedoms>.

²⁸⁶ *Ibidem*.

²⁸⁷ *Ibidem*.

²⁸⁸ Detectable cases are defined in Décret n° 2023-828 du 28 août 2023 relatif aux modalités de mise en œuvre des traitements algorithmiques sur les images collectées au moyen de systèmes de vidéoprotection et de caméras installées sur des aéronefs, pris en application de l'article 10 de la loi n° 2023-380 du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024 et portant diverses autres dispositions. Available at: <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000048007135>.

²⁸⁹ Décret n° 2023-939 du 11 octobre 2023 relatif aux modalités de pilotage et d'évaluation de l'expérimentation de traitements algorithmiques d'images légalement collectées au moyen de systèmes de vidéoprotection et de caméras installées sur des aéronefs. Available at: <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000048197679>.

²⁹⁰ See CNIL, *2024 Olympics: CNIL's Q&A on your privacy and freedoms*.

²⁹¹ The decision to use automated cameras for security purposes finds its motivation in the climate of terror that has characterised Paris over the last decade, due to the multiple terrorist attacks that have involved France and Paris specifically. See J. COAFFEE, *Evolving security motifs, Olympic spectacle and urban planning legacy*, cited, 650-651.

permanent surveillance system for public spaces in France²⁹². Supporters, however, maintained that these technologies were essential to uphold the constitutional objective of protecting public order and security, especially during events of significant scale. Ultimately, the matter was brought before the Conseil constitutionnel²⁹³, which ruled that the use of augmented cameras was consistent with the French Constitution²⁹⁴.

Indeed, under the introduced framework, images collected through authorised CCTV systems or drone-mounted cameras may be analysed solely for the purpose of detecting predefined events²⁹⁵ and issuing real-time alerts to the competent law enforcement authority. Alerts cannot, in themselves, form the basis for any individual decision or prosecutorial action. Algorithmic processing may only generate an attention alert strictly limited to indicating the predetermined event(s) it was programmed to detect, for the sole purpose of enabling appropriate follow-up measures. Among the safeguards laid down in the provision, the use of algorithmic processing is subject to a prior authorisation that must set out all relevant details of the processing and may be granted only where it is proportionate to the purpose pursued²⁹⁶. The public must also be informed in advance, by any appropriate means, of the use of algorithmic processing on the images collected²⁹⁷. Furthermore, the contested provisions expressly prohibit the use of face recognition, biometric identification and the processing of biometric data. The system may not carry out any automated cross-checking, interconnection or linkage with other personal data processing operations. In addition, the system must comply with a series of requirements that must remain verifiable throughout its operation. Where machine learning is used, the training, validation and test data must be relevant, adequate, representative and secure, and must include

²⁹² To cite a few, Amnesty International, *France: Intrusive Olympics surveillance technologies could usher in a dystopian future*, 20 March 2023. Available at: [France: Intrusive Olympics surveillance technologies could usher in a dystopian future - Amnesty International](#). L. KAYALI, *France sets EU precedent with 2024 Olympics surveillance arsenal*. Article published in *Politico*, 23 March 2023. Available at: [France sets EU precedent with 2024 Olympics surveillance arsenal – POLITICO](#). A. TOOMEY MCKENNA, *The Conversation Us, Paris Olympics Will Be a Training Ground for AI-Powered Mass Surveillance*. Article published on *Scientific American*, 24 July 2024. Available at: <https://www.scientificamerican.com/article/paris-olympics-will-be-a-training-ground-for-ai-powered-mass-surveillance/>. J. JABKHIRO, J. PRETOT, *How France plans to use AI surveillance to keep Paris 2024 Olympics safe*. Article published on *Reuters*, 8 March 2024. Available at: <https://www.reuters.com/sports/olympics-how-france-plans-use-ai-keep-paris-2024-safe-2024-03-08/>. La Quadrature du Net, *Algorithmic Video Surveillance*, cited, 54.

²⁹³ Conseil Constitutionnel, Décision no 2023-850 DC du 17 mai 2023, paragraph 26-49. Available at: https://www.legifrance.gouv.fr/download/pdf?id=yUQ_DLw0skm8dWcA3FHSXGOhrN4FcEKuXTAQRsVaHPI=.

²⁹⁴ The Constitutional Council concluded that LOI n° 2023-380 du 19 mai 2023 is consistent with the French Constitution, although the use of enhanced cameras must comply with the reservation mentioned in paragraph 39 of the decision.

²⁹⁵ The algorithmic processing of such images may be implemented only to ensure the security of sporting, recreational, or cultural events which, due to the scale of attendance or their circumstances, are particularly exposed to risks of terrorism or serious threats to personal safety. *Ibidem*, paragraph 37. The legislature could, without exceeding the scope of its competence, refer to a decree issued after consultation with the National Commission on Informatics and Liberties the task of specifying the predetermined events likely to present or reveal such risks and the characteristics of situations justifying the use of the processing.

²⁹⁶ The authorisation must be granted by the representative of the State in the department or, in Paris, by the Prefect of Police.

²⁹⁷ Unless circumstances prevent it or such information would undermine the objective pursued.

safeguards against bias. The system must also ensure automatic logging of flagged events for traceability, incorporate human oversight and risk-management mechanisms to prevent or correct bias or misuse, provide clearly defined procedures for immediate interruption, and undergo a validated test phase under conditions comparable to its authorised use. Where development is entrusted to a third party, that party must supply full technical documentation, demonstrate technical competence and service continuity, and disclose any relevant financial interests. CNIL and ANSSI exercise their respective oversight functions. Moreover, compliance with these requirements must be certified by the competent administrative authority, and the certificate must be published before the system can be made available to authorised services.

In Italy, the debate on AVS has been particularly wide-ranging, although the country is still in an experimental phase. It is interesting to note that it is mainly local authorities that are taking steps to introduce video surveillance systems augmented by AI. Two projects stand out as particularly illustrative: the ARGO project, piloted by the municipality of Turin, and the MARVEL project, piloted by the municipality of Trento. The main innovative element of ARGO was the possibility of enhancing video surveillance devices with distributed intelligence, which would have made it possible to extract data and metadata from images captured – e.g. type of vehicle, colour, number plate, country of registration, direction and speed, distinction between man/woman, colour of clothes and shoes, presence of objects such as bags, backpacks, hats – for follow-up operations. The main motivation for its introduction was that the use of computer vision algorithms would help the police in the fight against crime, both in the preventive and reactive fields. The project’s origins date back to around 2018²⁹⁸; following its approval in 2020²⁹⁹, it quickly became the subject of significant controversy. In 2021, an NGO – the Hermes Centre for Transparency and Digital Human Rights – requested the intervention of the Italian Data Protection Authority (hereinafter “DPA”), claiming that surveillance cameras installed throughout the city could identify and track an individual in public spaces and that Italy lacked

²⁹⁸ For the following discussion reference is made to E. D’ALBERGO, T. FASCIANI, G. GIOVANELLI, *La governance dell’Intelligenza Artificiale nelle politiche locali: trade-off e potere nel caso della videosorveglianza a Torino*, in *Rivista Trimestrale di Scienze dell’Amministrazione*, 4, 2023, 1-26. In 2016, the City of Turin won a grant from the Italian government to redevelop the suburbs. The project was called “*AxTO azioni per le periferie torinesi*”. Available at: http://www.comune.torino.it/arredourbano/bm~doc/allegato_2_schede-descrittive-azioni.pdf. One of the project’s actions was the implementation of an integrated video surveillance system for urban security purposes (5.1). The project was approved based on the proposition of the Mayor of Turin and the scheme made by the local police department of Turin (Reparto Investigativo Tecnologico) and by the Società 5T s.r.l. in 2018. In 2019 an agreement between the municipality of Turin, the Region of Piemonte, the police department (Prefettura di Torino) and other non-institutional authors was concluded. The agreement required an improvement of video surveillance devices as a priority (Art. 1). See *Accordo per la Sicurezza Integrata e lo Sviluppo della Città di Torino*. Available at: https://www.interno.gov.it/sites/default/files/allegati/protocollo_per_la_sicurezza_integrata_e_lo_sviluppo_torino.pdf. See also M. MAZZARELLA, *Argo: un nuovo progetto di videosorveglianza per le strade di Torino*, published in *Osservatorio sullo Stato Digitale*, IRPA, 9 November 2021.

²⁹⁹ Comune di Torino, Delibera n. 2020 01738/048, 4 August 2020. Available at: http://www.comune.torino.it/giunta_comune/intracom/htdocs/2020/2020_01738.pdf.

a legal basis for such use³⁰⁰. Following this report, the DPA launched an investigation by requesting information from the municipality of Turin³⁰¹. The municipality replied that the project was still in a preliminary stage and that decisions on data protection had yet to be taken. In September 2021, a parliamentary interpellation was submitted to the Government requesting information on the use of video surveillance cameras enhanced with face recognition technology in the national territory³⁰². In providing an answer, the Undersecretary of State for the Interior affirmed that ARGO, which was still in a preliminary stage, did not involve the processing of biometric data³⁰³. However, the controversy surrounding the project did not subside and GDPR compliance proved particularly costly. Therefore, although the DPA had not inhibited the use of intelligent video surveillance devices and had rather initiated the process to bring it into compliance with data protection legislation, the municipality assessed the revisited project as not adequate for the time and financial resources allocated to the project. Therefore, the project was later partially revised, not implementing the computer vision algorithms initially planned³⁰⁴.

The difficulty in ensuring compliance with the legal framework is clearly shown by the European-funded project Multimodal Extreme Scale Data Analytics for Smart Cities Environments (hereinafter “MARVEL”)³⁰⁵. MARVEL aims to combine different types of data, various technologies and data analysis to leverage data produced daily in smart city environments in order to support local public authorities in addressing multiple social challenges³⁰⁶. MARVEL’s prerogative in the specific case of the municipality of Trento, which was chosen for a real-life experiment, was to increase urban security through the analysis of audio and visual data, searching for unusual occurrences (e.g. crowded areas,

³⁰⁰ Hermes Centre, *Segnalazione al Garante Privacy — Progetto ARGO Torino*, 2021. Inquiry based on art. 57 paragraph 1 GDPR.

³⁰¹ Italian Data Protection Authority, *Telecamere intelligenti a Torino, il Garante privacy apre un’istruttoria. Inviata una richiesta di informazioni al Comune*. Available at: <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/10035760>.

³⁰² Interpellanza parlamentare, *Iniziative di competenza volte a tutelare i diritti costituzionali dei cittadini in relazione all'utilizzo di impianti di videosorveglianza con sistema di riconoscimento facciale – 2-01330*, Annex A). Available at: https://www.camera.it/leg18/410?idSeduta=0571&tipo=documenti_seduta&pag=allegato_a#.

³⁰³ The full answer is available at: <https://www.camera.it/leg18/410?idSeduta=0571&tipo=stenografico>.

³⁰⁴ This change was reported by E. D’ALBERGO, T. FASCIANI, G. GIOVANELLI, *La governance dell’Intelligenza Artificiale nelle politiche locali*, cited, 15. As the authors report, the financial resources required for the development of the project in compliance with the GDPR were too high. Therefore, the municipality decided to discontinue the development and implementation of the project, using the financial resources allocated for it to improve the surveillance system, without enhancing it with AI. See Consiglio Comunale Città di Torino, *Risposta all’Interpellanza 2023-00105*. Available at: https://servizi.comune.torino.it/consiglio/prg/documenti/documentazione/elenchi_atti/ElencoInterpellanze.html.

³⁰⁵ During my doctoral studies, I carried out part of my research at the General Directorate of the municipality of Trento, which was involved, among other things, in the MARVEL project at that time.

³⁰⁶ Information about the MARVEL Project can be found on the website <https://marvel-project.eu/>. See also D. BAJOVIC ET AL., *MARVEL: Multimodal Extreme Scale Data Analytics for Smart Cities Environments*, 2021 *International Balkan Conference on Communications and Networking (BalkanCom)*, Novi Sad, Serbia, 2021, 143-147. Available at: doi: 10.1109/BalkanCom53780.2021.9593258.

criminal or antisocial behaviour), and consequently alerting competent authorities to intervene promptly.

The project's documents clearly state that MARVEL was developed with the precise aim to preserve the privacy of natural persons. To this end, although in simplified terms³⁰⁷, MARVEL is based on a three-layer design³⁰⁸ and uses multiple technologies. First, it uses AI algorithms (machine learning and deep learning) to learn from the data. For example, it trains models on city video and audio so the system can recognise unusual events, such as a loud shout. Moreover, to harness information contained in the dataset while respecting privacy, it uses federated learning: instead of sending all raw data to one place, MARVEL trains on data locally³⁰⁹. This lets the system improve without moving personal data around³¹⁰. Moreover, in order to promote privacy by design, data are automatically anonymised at the edge (*i.e.* at the source)³¹¹. For example, a technique is to blur or swap faces captured by a camera or convert voices captured by a microphone, so that personal data are “removed” in the first processing stage³¹². Although MARVEL was developed with the precise goal of ensuring the protection of citizens' privacy, the project has received much criticism from both political parties and citizens³¹³.

³⁰⁷ For a proper technical discussion see D. BAJOVIC ET AL., *MARVEL: Multimodal Extreme Scale Data Analytics for Smart Cities Environments*.

³⁰⁸ It is based on an Edge-Fog-Cloud (E2F2C) architecture. This means that three steps of processing can be identified. First, data is captured and processed at the edge, by the local device itself (camera or microphone). The edge devices might run real-time checks, quick and simple analysis (for example, quickly detecting a moving car). If more computing is needed, the data goes up to fog and finally to cloud. The cloud performs deep analysis, and advanced AI train models on the collected data to recognise complex patterns (e.g. recognise a pedestrian or a siren sound), and then new knowledge (updated AI models) is sent back down to the edge devices for use. This layered setup lets MARVEL react quickly to unusual events (using edge or fog) while also learning from long-term data in the cloud.

³⁰⁹ Therefore, data does not leave the proximity of the sensor that generated and collected them. Each device or local server can update an AI model with its own data, and only the updates (not the raw video or sound) are shared and merged. See Deliverable D3.4: *MARVEL's federated learning realization*. Available at: <https://marvel-project.eu/deliverables/>.

³¹⁰ As far as security is concerned, all information sent between devices, local servers, and the cloud is encrypted. In other words, the data is scrambled during transmission so outsiders cannot easily read. Edge devices in MARVEL include security features (such as special chips) to prevent hacking or tampering. Combined with encrypted links, this «bulletproofs» the system so that data and devices stay safe.

³¹¹ For a specific description of the mechanisms deployed in MARVEL to preserve data privacy in the Edge-to-Fog-to-Cloud (E2F2C) see Deliverable D3.3: *E2F2C Privacy preservation mechanisms*. Available at: <https://marvel-project.eu/deliverables/>.

³¹² See R. CHEN, X. CHEN, B. NI, Y. GE, *Simswap: An efficient framework for high fidelity face swapping*, in *ACM International Conference on Multimedia*, 2020, 2003–2011. Available: <https://doi.org/10.1145/3394171.3413630>.

The importance of the implementation of anonymisation and pseudonymization techniques was specifically stressed by the ethics evaluation, since it enables to decrease the possibility of any harm potentially caused by data processing. More information about the ethical evaluation of projects funded by the European Union can be found at: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/ethics_en.htm. As specified in project's documents, the «data are not used for the profiling of citizens» and MARVEL was «not able to analyse or understand conversations, but only to associate audio and video with situations considered dangerous or anomalous». MARVEL, Deliverable D3.3, cited, 18.

³¹³ Il “Grande fratello” in cinque piazze, fra microfoni e telecamere: Trento laboratorio per tre progetti. Article published on *L'Adige.it*, 9 May 2023. Available at: <https://www.ladige.it/cronaca/2023/05/09/il-grande-fratello-in-cinque-piazze-fra-microfoni-e-telecamere-trento-laboratorio-per-tre-progetti-1.3493158>. Smart city o città della sorveglianza? L'esemplare caso di Trento. Article published on *L'Indipendente*, 22 November 2024. Available at: <https://www.lindipendente.online/2024/11/22/smart-city-o-citta-della-sorveglianza-lesemplare-caso-di-trento/>.

This led the DPA to start an investigation, asking the municipality of Trento for information about the project. In the end, the DPA stated that the processing of personal data was not compliant with the GDPR, blocked the project and sanctioned the municipality of Trento³¹⁴.

Neither of the projects reached a conclusion, albeit for different reasons. In both cases, the primary issue lay in their failure to comply with data protection legislation, as outlined by the DPA. However, in both cases there was a significant lack of clarity regarding the actual functionality of the cameras used in the projects. This lack of clarity, mainly due to a lack of transparency about the projects, led to the discontinuation of both, with negative repercussions on public trust.

Most contemporary video-analytics systems described as “intelligent” fall within the AI Act’s definition of an AI system. This is the case whenever the technology infers information from video data, rather than merely applying simple, fixed rules. For example, automated surveillance may be considered an AI system when cameras are equipped with tools capable of recognising suspicious behaviour, detecting crowd movements or loitering, identifying fights or abandoned objects, signalling intrusions into restricted areas, performing face recognition, categorising individuals based on facial or bodily features, tracking and re-identifying persons, or detecting recurrent behavioural patterns and generating corresponding alerts. Only very simple rule-based systems – such as basic motion detection or fixed sensor-triggered alarms – seem to fall outside this definition.

1.2. Identifying individuals through biometric technologies

In the previous paragraph I analysed AVS in general. In this paragraph, I will focus on biometric technologies, since the AI Act specifically deals with them in relation to prohibited AI practices and high-risk AI systems. In order to properly understand the field that the law regulates and to properly address legal considerations in the following Chapters, I will provide some technical details on biometric technologies.

Police departments have always invested in systems for the identification of individuals, since accurate identification is fundamental for an effective and efficient action in the context of law enforcement. From early methods, such as fingerprinting, to modern tools like DNA analysis, biometric technologies have always been the focus of attention, significantly enhancing the capability of law enforcement

³¹⁴ Garante della protezione dei dati personali, *Decision 11 January, No. 5*. Available at: <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9977020>.

authorities to reliably link a person (a suspect, a victim, a witness) to specific objects (vehicles, weapons, stolen goods), locations and events³¹⁵.

Biometric technologies involve the specific processing of data related to a natural person's physical, physiological, behavioural, or psychological characteristics in order to enable or confirm his or her unique identification³¹⁶. First and foremost, a distinction between first-generation biometric technologies and second-generation biometric technologies can be identified³¹⁷. The former focuses on “strong identifiers” that allow the unique identification of a natural person, such as fingerprints, irises and faces. One of the most discussed uses of strong biometric technologies in law enforcement is face recognition technology³¹⁸. Conversely, second-generation technologies focus on “weak identifiers”, such as body shape or behavioural patterns. This type of biometrics is relevant in the context of law enforcement, as it might allow to identify subjects with abnormal or suspicious behaviour³¹⁹. A different type of identifiers in second-generation biometric technologies are “soft identifiers”, which focus on more generic features, such as gender, age or ethnic origins, which are not uniquely associated to an individual³²⁰.

To be considered a biometric technology, therefore, two conditions must be fulfilled: (i) the data processed must be related to physical, physiological, behavioural or psychological characteristics of a

³¹⁵ G. MOBILIO, *Your face is not new to me – Regulating the surveillance power of facial recognition technologies*, in *Internet Policy Review*, 12(1), 2023, 2.

³¹⁶ The NIST defines «biometrics» in the following way: «In information technology, biometrics usually refers to automated technologies for authenticating and verifying human body characteristics such as fingerprints, eye retinas and irises, voice patterns, facial patterns, and hand measurements». NIST Glossary, Term «Biometrics». Accessible at: <https://csrc.nist.gov/glossary/term/biometrics>. C. WENDEHORST, Y. DULLER, *Biometric Recognition and Behavioural Detection. Assessing the ethical aspects of biometric recognition and behavioural detection techniques with a focus on their current and future use in public spaces*, Study commissioned by the European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs, August 2021, 12. For an interesting study on biometric technologies see A. KAK (Ed.), *Regulating Biometrics: Global Approaches and Urgent Questions*, AI Now Institute, September 2020. See also Consultative Committee Convention 108, *Facial Recognition: current situation and challenges*, cited.

³¹⁷ E. MORDINI, D. TZOVARAS (Eds.), *Second Generation Biometrics: The Ethical, Legal and Social Context*, Springer, 2012.

³¹⁸ FRT is currently being tested and used in many different contexts and for different purposes, both by private and public actors. S. FELDSTEIN, *The Global Expansion of AI Surveillance*, Working Paper, Carnegie Endowment for International Peace, September 2019. Available at: <https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847>. The author stresses that 43 percent of the States (precisely 64 out of a reference sample of 176 States) used facial recognition technologies for surveillance purposes. East and Pacific Asia are the areas where there has been the most widespread use of FRT. By contrast, in Europe and Eurasia less than a half of the considered Countries have adopted FRT. An overview of the current uses of FRT around the world can be seen in S. DEL GATTO, *Il riconoscimento facciale. A che punto siamo?*, in *Giornale di diritto amministrativo*, 5, 2022, 692-699. Examples of usage by private companies, consumers and governments can be found in T. MADIEGA, H. MILDEBRAT, *Regulating facial recognition in the EU*, European Parliamentary Research Service, September 2021, 3-4. During the Covid-19 pandemic FRT was used vastly. See M. VAN NATTA ET AL., *The rise and regulation of thermal facial recognition technology during the COVID-19 pandemic*, in *Journal of Law and the Biosciences*, 7(1), 2020, Article Isaa038.

³¹⁹ C. WENDEHORST, Y. DULLER, *Biometric Recognition and Behavioural Detection*, cited, 14.

³²⁰ A. DANTCHEVA, P. ELIA, A. ROSS, *What else does your biometric data reveal? A survey on soft biometrics*, in *IEEE Transactions on Information Forensics and Security*, 11(3), 2016, 441-467. B. HASSAN, E. IZQUIERDO, T. PIATRIK, *Soft biometrics: a survey. Benchmark analysis, open challenges and recommendations*, in *Multimedia Tools and Applications*, 83, 2024, 15151-15194.

natural person and (ii) the data processing must be aimed at identifying the subject. Technological advances (e.g. enhanced sensor and computer capabilities, high-resolution cameras, IoT) and improvements in AI and data analysis have greatly increased the potential of biometric technologies, giving rise to ethical and legal concerns, especially when used by public authorities in public spaces³²¹.

2. An example of hard biometrics: face recognition technology

One of the most discussed uses of (strong) biometric technologies in law enforcement is face recognition technology (hereinafter “FRT”)³²². A further reason to focus on FRT lies in the fact that the AI Act, while not explicitly addressing it, appears to have FRT in mind when laying down its provisions on biometric identification.

Generally, the use of FRT for law enforcement purposes involves the implementation of this software functionality into existing devices (e.g. image databases, CCTVs, body cameras, drones). A critical use of FRT is its integration with CCTV cameras in public spaces, especially when deployed in real-time, commonly referred to as “live FRT”³²³. Public spaces, such as streets and train stations, are by nature accessible to everyone without restrictions³²⁴, and individuals in these settings generally have a reasonable expectation of remaining anonymous³²⁵. However, with FRT-enabled CCTV cameras, people’s facial images can be captured and analysed easily, often without their knowledge or consent³²⁶. It is precisely the ability to detect faces from a considerable distance without requiring the

³²¹ A growing interest in the use of FRT in public spaces was described by K. CRAWFORD ET AL., *AI Now Report*, 2020, 11.

³²² G. GONZÁLEZ FUSTER, *Artificial Intelligence and Law Enforcement: Impact on Fundamental Rights*, Study commissioned by the European Parliament, 2020.

³²³ In real-time biometric identification the monitoring is constant and the number of people involved by the processing of biometric data is indeterminate. A.A. GIKAY, *Regulating use by law enforcement authorities of live facial recognition technology in public spaces: an incremental approach*, in *Cambridge Law Journal*, 82(3), 2023, 414,-449.

³²⁴ For the purpose of this Thesis, I use the definition encapsulated in Article 3(44) AI Act « ‘publicly accessible space’ means any publicly or privately owned physical place accessible to an undetermined number of natural persons, regardless of whether certain conditions for access may apply, and regardless of the potential capacity restrictions » See also Recital 19.

³²⁵ See Article 29 Working Party, *Opinion 3/2012 on developments in biometric technologies*, April 2012, 2. J. LAPERRUQUE, *Facing the Future of Surveillance*, The Constitution Project’s Task Force on Facial Recognition Surveillance, Project on Government Oversight, 25. The expectation of maintaining anonymity in public spaces is fundamental in democratic societies, as it is a prerequisite for exercising one’s rights in a truly free manner. For an interesting contribution on privacy in public spaces see J.R. REIDENBERG, *Privacy in Public*, in *University of Miami Law Review*, 69(1), 2014, 141-159. On the right to anonymity see G. FINOCCHIARO, *Anonimato*, in *Dig. Disc. Priv.*, sez. Civ., Aggiornamento V, 2010, 13. G. FINOCCHIARO (Ed.), *Diritto all’anonimato. Anonimato, nome e identità personale*, CEDAM, 2008.

³²⁶ FRA, *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, Publications Office of the European Union, 2020, 24. It is noteworthy that citizens could have different expectations about spaces in which FRT is carried out. For example, in border crossing controls the use of FRT may be expected and considered more acceptable than in other public spaces, such as streets.

subject's awareness or cooperation that makes FRT a particularly attractive biometric technology in the context of law enforcement³²⁷.

Although any use of FRT requires attention, the use of such technologies for the prevention, investigation and prosecution of crimes and the execution of criminal penalties, which is undoubtedly a high-stake operational environment, requires specific consideration³²⁸. Using FRT in the context of law enforcement is certainly promising, because of the benefits for public security³²⁹, but it is also worrisome³³⁰, as it poses serious threats to fundamental rights and freedoms³³¹. It becomes thus necessary to strike a delicate balance between security and public order, on the one hand, and individual rights and freedoms, on the other. Striking a balance requires to take into account multiple aspects: the purpose for which FRT is used (e.g. for investigative purposes, for preventive purposes), the modality of its usage (e.g. real-time or delayed mode)³³², the context of its usage (e.g. public spaces, airports, borders) and the number of people targeted (e.g. one, many, indeterminate)³³³.

³²⁷ Especially if FRT is compared to iris scanning or fingerprinting, which require the knowledge and cooperation of the subject. However, in the law enforcement sector, subjects of interest are likely to be uncooperative. See L. INTRONA, H. NISSEMBAUM, *Facial Recognition Technology A Survey of Policy and Implementation Issues*, Lancaster University Management School Working Paper 2010/030, 10.

³²⁸ Even when FRT is not involved, in relation to surveillance in general, the Italian Data Protection Authority stressed the need of strict requirements to prevent video surveillance expanding to the point of restricting citizens' rights, prescribing an essential balance between citizens' rights and freedoms, on the one hand, and security and crime prevention, on the other hand. Garante per la protezione dei dati personali, *Video Surveillance*. Decision dated 8 April 2010. Available at: <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/1734653>. See also E. SCHNEIDER, "Captazioni" fuori legge: che cosa fare per garantire una effettiva tutela dei diritti dei sorvegliati?, published in *Osservatorio sullo Stato Digitale*, IRPA, November 10, 2022.

³²⁹ FRT may be used to identify individuals suspected of having committed a crime, to search for missing people, for border control or migration management, or it may be employed in stadiums, during concerts, or at public events to maintain public security and order. See A.A. GIKAY, *Regulating the Use of Live Facial Recognition Technology*, cited, 6.

³³⁰ People seem more likely to trust law enforcement authorities to use FRT responsibly than advertisers or tech companies. See A. SMITH, *More Than Half of U.S. Adults Trust Law Enforcement to Use Facial Recognition Responsibly*, published in *Pew Research Centre*, 5 September 2019. See also S. DEL GATTO, *Riconoscimento facciale e uso dei servizi governativi. Numerosi benefici, ma quanti i rischi?*, published in *Osservatorio sullo Stato Digitale*, IRPA, 21 January 2021.

³³¹ Examples of rights and freedoms that could be negatively impacted are right to privacy and right to protection of personal data, right to dignity, equality, freedom from discrimination, freedom of assembly, association and expression, right to an effective remedy and to a fair trial and raise ethical questions. FRA, *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, cited; United Nations High Commissioner for Human Rights, *Impact of new technologies on the promotion and protection of human rights in the context of assemblies, including peaceful protests*, Report, 2020. For a detailed discussion on the risks posed by FRT to fundamental rights see G. MOBILIO, *Tecnologie di riconoscimento facciale. Rischi per i diritti fondamentali e sfide regolatorie*, Editoriale Scientifica, 2021. Even religious freedoms are threatened by the use of FRT. Among others, see E.A. ROWE, *Regulating Facial Recognition Technology in the Private Sector*, in *Stanford Technology Law Review*, 24(1), 2021, 31.

³³² As Giuseppe Mobilio recognises the use of FRT by law enforcement authorities represents «a scenario where risks and benefits come closer together». G. MOBILIO, *Your face is not new to me*, cited, 3.

³³³ As I will discuss in more detail in Chapter 3, I agree with the approach followed by those authors who do not advocate a general ban on the use of facial recognition, but rather look at the specific purpose, modalities and contexts of its use, and carefully assess the risks that the specific use of FRT raises for fundamental rights and freedoms and, finally, identify limits and conditions that allow FRT to be used in a way that is in accordance with constitutional principles and public values. Among others, G. MOBILIO, *Your face is not new to me*, cited; G. MOBILIO, *Tecnologie di riconoscimento facciale*, cited. S. NAKAR, D. GREENBAUM, *Now you see me. Now you still do: facial recognition technology and the growing lack of privacy*, in *Journal of Science & Technology Law*, 23(1), 2017, 88.

2.1.Face recognition technology: how does it work?

In recent years, FRT has undergone a major evolutionary process, which inevitably shapes our understanding of this technology and our evaluation of it³³⁴. The origins of FRT can be traced back to the mid twentieth century, when Woodrow Wilson Bledsoe, an American mathematician and computer scientist, first attempted to recognise faces in a computational form³³⁵. The method used, which consisted of encoding each individual with a vector of calculated distances between facial features³³⁶, required the presence of a human operator to identify landmarks in a facial image and enter them into the machine. This method was computationally expensive and slow. Some years later, a method to automatically extract such landmarks from facial images was first proposed by Takeo Kanade in his Doctoral Thesis³³⁷, however, the technology used to recognise faces in an image was still based on the geometrical structure of the face³³⁸. Despite face recognition gaining more attention in the 1970s, the limits of computing power and lack of sufficient data did not allow for particularly noticeable improvements.

The following decades (1980s and 1990s) saw considerable improvements in FRT, especially due to the development of different methods for automatically recognising faces, such as the Principal Component Analysis (PCA)³³⁹, which allowed facial features to be extracted from pixels in the whole image, instead of relying on face geometry for recognition³⁴⁰. Later in the 1990s, other face recognition approaches were developed³⁴¹. In addition to better technological capabilities, the U.S. Department of

³³⁴ The literature on the topic is huge. Among others, see W. ZHAO, R. CHELLAPPA, P.J. PHILLIPS, A. ROSENFELD, *Face recognition: A literature survey*, in *ACM computing surveys*, 35(4), 2003, 399–458; M. SHARIF, F. NAZ, M. YASMIN, M.A. SHAHID, A. REHMAN, *Face recognition: A survey*, in *Journal of Engineering Science & Technology Review*, 10(2), 2017, 166-177.

³³⁵ W.W. BLEDSOE, *The model method in facial recognition*, Panoramic Research Inc., Palo Alto, CA, Rep. PR1, 15(47), 1966, 2. W.W. BLEDSOE, *Some Results on Multicategory Pattern Recognition*, in *Journal of the ACM*, 13 (2), 1966, 304–316.

³³⁶ I.D. RAJI, G. FRIED, *About Face: A Survey of Facial Recognition Evaluation*, 2021. Available at: <https://arxiv.org/pdf/2102.00813>.

³³⁷ T. KANADE, *Picture Processing System by Computer Complex and Recognition of Human Faces*, PhD Thesis, 1974. Available at: https://www.ri.cmu.edu/pub_files/pub3/kanade_takeo_1973_1/kanade_takeo_1973_1.pdf.

³³⁸ L. LI, X. MU, S. LI, H. PENG, *A Review of Face Recognition Technology*, in *IEEE Access*, 8, 2020, 139110-139120. Available at: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9145558>.

³³⁹ For a discussion on Principal Component Analysis (PCA) see L. QINJUN, C. TIANWEI, Z. YAN, W. YUYING, *Facial Recognition Technology: A Comprehensive Overview*, in *Academic Journal of Computing & Information Science*, 6(7), 2023, 15-26. In the 1990s the PCA method was further developed. M. TURK, A. PENTLAND, *Eigenfaces for recognition*, in *Journal of cognitive neuroscience*, 3(1), 1991, 71-86. J. LI, B. ZHAO, H. ZHANG, J. JIAO, *Face recognition system using SVM classifier and feature extraction by PCA and LDA combination*, 2009 International Conference on Computational Intelligence and Software Engineering, Wuhan, China, 2009, 1-4. Available at: <https://ieeexplore.ieee.org/document/5364125>.

³⁴⁰ For a technical discussion on the evolution of different techniques see National Academies of Science, Engineering and Medicine, *Facial Recognition Technology: Current Capabilities, Future Prospects, and Governance*, Washington, DC, The National Academies Press, 34.

³⁴¹ The reference goes to Linear Discriminate Analysis (LDA). Among others, see P.S. PENEV, J.J. ATICK, *Local feature analysis: a general statistical theory for object representation*, in *Network: Computation in Neural Systems*, 7(3), 1996,

Defense and the National Institute of Standards and Technology (NIST) funded the Face Recognition Technology (FERET) database³⁴², providing researchers with the necessary data to make progress in this field³⁴³. Moreover, academic research interest was promoted by competitions and investments, especially from governments which supported advances in FRT mainly for law enforcement purposes, to enable criminal surveillance and criminal investigations³⁴⁴.

Advances in the technological sphere, particularly in the field of AI, and the amount of data available have further supported improvements in FRT. In the 2000s, methods developed within computer vision – such as simple neural networks – started to be applied to face recognition³⁴⁵. In addition, large databases of facial images – such as the Labeled Faces in the Wild (LFW)³⁴⁶ – were created to support research in the field. Some years later, deep neural networks, precisely Deep Convolutional Neural Networks (DCNN), proved to be a breakthrough in image recognition, demonstrating enormous capabilities which became fully evident in the 2010s³⁴⁷. The impressive performance of deep learning in face recognition stems from its ability to automatically learn how to extract relevant facial features during the training process, eliminating the need for manual feature engineering: instead of relying on handcrafted features, deep learning models are trained on large datasets of facial images, captured under both controlled and uncontrolled conditions. Through this data-driven process, DCNNs automatically extract features from facial images and compares extracted features. As a result, DCNN-based approaches have become the foundation of modern FRT systems and continue to evolve as a mainstream method in the field³⁴⁸. The increased complexity of DCNNs, however, leads to reduced

477–500; P. N. BELHUMEUR, J. P. HESPANHA, D. J. KRIEGMAN, *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection*, in *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 19(7), 1997, 711-720. See also Y. MING-HSUAN, D. J. KRIEGMAN, N. AHUJA, *Detecting Faces in Images: A Survey*, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1), 2002, 34.

³⁴² Information about FERET are available at: <https://www.nist.gov/programs-projects/face-recognition-technology-feret>.

³⁴³ I. D. RAJI, G. FRIED, *About Face*, cited. The dataset was created by a collection of facial images in a semi-controlled environment. Therefore, the quality of face images was quite high. P. PHILLIPS, H. MOON, S. RIZVI, P. RAUSS, *The FERET evaluation methodology for face-recognition algorithms*, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22, 1090–1104.

³⁴⁴ P. J. PHILLIPS, P. J. FLYNN, T. SCRUGGS, K. W. BOWYER, J. CHANG, K. HOFFMAN, J. MARQUES, J. MIN, W. WOREK, *Overview of the face recognition grand challenge*, in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR '05)*, 1, 2005, 947–954.

³⁴⁵ For a discussion on different approaches to find the ideal transformation function so as to achieve the optimal recognition effect see L. LI, X. MU, S. LI, H. PENG, *A Review of Face Recognition Technology*, cited.

³⁴⁶ Contrary to FERET dataset, Labeled Faces in the Wild (LFW) contains unconstrained face images in an infinite combination of poses, illumination conditions, and expressions.

³⁴⁷ The reference necessarily goes to DeepFace, created by a research group at Facebook. See Y. TAIGMAN, M. YANG, M. RANZATO, L. WOLF, *DeepFace: Closing the Gap to Human-Level Performance in Face Verification*, 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, 1701-1708. M. WANG, W. DENG, *Deep Face Recognition: A Survey*, 2020. Available at: arXiv:1804.06655.

³⁴⁸ For a detailed discussion on the functioning of a DCNNs see National Academies, *Facial Recognition Technology*, cited, 36-38.

explainability of the processes they use to generate outputs from given inputs, making their decision-making less transparent and harder for humans to comprehend.

From a technical standpoint, FRT is a subfield of the broader domain of pattern recognition³⁴⁹. Accordingly, FRT is a probabilistic technology³⁵⁰, meaning its outcomes are best understood as predictions, as defined in Chapter 1³⁵¹. Interestingly, a definition of FRT has not been included in the data protection legislation, neither in the GDPR nor in the LED. As for the AI Act, it does not define FRT either, rather focusing on the term “biometric identification”³⁵². This term is narrower than “face recognition” in the sense that it covers only one specific application of FRT, namely face identification but not face verification. However, at the same time, it is broader because it potentially includes biometric techniques that use elements other than faces, such as gaits, irises, fingerprints or voices, to recognise the individual. In 2012, Article 29 Working Party – the European Data Protection Board’s predecessor – specifically defined FRT in general terms as the “*automatic processing of digital images which contain the faces of individuals for identification, authentication/verification or categorisation of those individuals*”³⁵³. As I will analyse shortly, this definition is still relevant as it identifies the different tasks for which FRT can be used, moving away from the “one-size-fits-all approach” that is misleading in this field.

³⁴⁹ INTRONA, NISSENBAUM, *Facial Recognition Technology*, cited, 10, where facial recognition is described as «*a subdivision problem of visual pattern recognition*». Pattern recognition technology «*uses statistical techniques to detect and extract patterns from data in order to match it with patterns stored in a database*». They define an automated biometric system essentially as «*a pattern recognition system that operates by acquiring biometric data (a face image) from an individual, extracting certain features (defined as mathematical artifacts) from the acquired data, and comparing this feature set against the biometric template (or representation) of features already acquired in a database*». *Ibidem*, 9. See also LI, MU, LI, PENG, *A Review of Face Recognition Technology*, cited, 139110.

³⁵⁰ In 2019 the CNIL defined FRT as «*a probabilistic software application that can automatically recognise a person based on their facial attributes in order to authenticate or identify them [...]*». CNIL, *Facial Recognition: For a debate living up to the challenges*, cited, 3. «*The facial recognition techniques used are based on an estimated match between templates: the one being compared and the baseline(s). From this point of view, they are probabilistic. From the comparison, a higher or lower probability is deduced that the person is indeed the person to be authenticated or identified*». *Ibidem*. The probabilistic nature of FRT is highlighted also by MOBILIO, *Tecnologie di riconoscimento facciale*, cited, 35.

³⁵¹ G. SARTOR, *L'informatica giuridica e le tecnologie dell'informazione*. Corso di informatica giuridica, Giappichelli, 2022, 315. FRT «*do not provide for a definitive result but rely on probabilities that two faces, or images of faces, correspond to the same person*». EDPB, *Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement*, April 2023, 13.

³⁵² AI Act, Article 3(35) «*‘biometric identification’ means the automated recognition of physical, physiological, behavioural, or psychological human features for the purpose of establishing the identity of a natural person by comparing biometric data of that individual to biometric data of individuals stored in a database*». See also Recital 15.

³⁵³ Article 29 Data Protection Working Party, *Opinion 02/2012 on facial recognition in online and mobile services*, March 2012, 2.

Face recognition could be described as a five-step process³⁵⁴. The first step is “image acquisition” in a digital format (so-called “probe image”³⁵⁵). Image acquisition is a crucial phase since the performance of FRT highly depends on its quality. The probe image might be acquired through a photograph or a video, under controlled or uncontrolled circumstances, voluntarily or involuntarily (*i.e.* with or without the knowledge and the cooperation of the subject). The second step is “face detection”, which implies a scanning of the probe image to detect the presence of face(s), identifying and separating the face from the surrounding background. This task is not easy, especially when the probe image is taken under uncontrolled circumstances and the face is surrounded by a complex background. The third step is “normalisation”, which represents the preparatory phase for feature extraction. Normalisation or standardisation requires rotating, centering and resizing the probe image and correcting pose orientation and illumination in order to make the facial landmarks evident. The fourth step is “feature extraction”, whose precise aim is to generate a biometric template of the probe image³⁵⁶. A template is a mathematical representation of the probe image, which is unique and specific to each individual and tends to persist over time³⁵⁷. Applying the concepts I analysed in Chapter 1, a trained model (e.g. a DCNN), generally based on a supervised learning process³⁵⁸, is used to convert a face image into a biometric template. The model is trained to extract from the face image only those features that are considered relevant for comparisons, ignoring the so-called “nuisance variations”³⁵⁹. Once a template

³⁵⁴ For this discussion reference is mainly made to National Academies, *Facial Recognition Technology*, cited. In other documents, facial recognition is presented as a two-step process. «‘Facial recognition’ is therefore a two-step process: the collection of the facial image and its transformation into a template, followed by the recognition of this face by comparing the corresponding template with one or more other templates». EDPB, *Guidelines 05/2022*, cited, 9. In other contributions, FRT is described as a four-step process. See INTRONA, NISSENBAUM, *Facial Recognition Technology*, cited, 15. In other cases it is presented as a six-step process. See MOBILIO, *Tecnologie di riconoscimento facciale*, cited, 32-33.

³⁵⁵ INTRONA, NISSENBAUM, *Facial Recognition Technology*, cited, 11.

³⁵⁶ A «biometric template» is a mathematical representation obtained by feature extraction from biometric data limited to the characteristics necessary to perform identification and verification. See Article 4(12) of Regulation (EU) 2019/818 of the European Parliament and of the Council of 20 May 2019 on establishing a framework for interoperability between EU information systems in the field of police and judicial cooperation, asylum and migration and amending Regulations (EU) 2018/1726, (EU) 2018/1862 and (EU) 2019/816. From a technical perspective, a «biometric template» is defined as a «set of stored biometric features comparable directly to biometric features of a recognition biometric sample». ISO/IEC 19792:2009 Information technology — Security techniques — Security evaluation of biometrics, 4.2.8.

³⁵⁷ EDPB, *Guidelines 05/2022*, cited, 9. Consultative Committee Convention 108, *Guidelines on facial recognition*, June 2021, 9 footnote 13. Initially, it was believed that biometric templates – especially fingerprint templates – were anonymous data, assuming that it was not possible to re-identify individuals to whom the fingerprints belonged. However, research has shown that it is possible, to a certain extent, to reconstruct the image of a fingerprint from a fingerprint template. See C. JASSERAND-BREEMAN, *Reprocessing of biometric data for law enforcement purposes: Individuals’ safeguards caught at the Interface between the GDPR and the ‘Police’ directive?*, PhD Thesis, University of Groningen, 15 March 2019. J. GRIJPINK, *Privacy Law: Biometrics and Privacy*, in *Computer Law & Security Report*, 17(3), 2001, 154-160. K. CAO, A. K. JAIN, *Learning Fingerprint Reconstruction: From Minutiae to Image*, in *IEEE Transactions on Information Forensics and Security*, 2015, 10(1), 104-117.

³⁵⁸ National Academies, *Facial Recognition Technology*, cited, 27. Supervised learning means that labelled samples are used for training the model.

³⁵⁹ National Academies, *Facial Recognition Technology*, cited, 33. Avoiding the nuisance variations is necessary to create a sufficiently distinctive biometric template, that is a FRT system that «produce[s] templates that when compared yield high similarity scores from the photos of the same persona and low scores for photos of different people». *Ibidem*.

has been generated, the fifth step is recognition or “comparison”³⁶⁰ or “face matching”³⁶¹, where a similarity score is generated, namely a numerical value that represents how similar two templates are to each other. A similarity score, rather than a binary result (match/non-match), is generally preferred in the context of FRT, as two face images are unlikely to be perfectly identical³⁶².

The system owner sets a numerical threshold against which the similarity score is evaluated³⁶³: similarity scores falling below this threshold are considered “mismatches”, while those exceeding it are treated as “matches”. The determination of this threshold is not a neutral or purely technical decision; rather, it reflects a value judgment that weighs the trade-offs between false matches and their associated costs. Once a match is identified, a human operator generally verifies it and decides what action to take.



Figure 10: The five-step process of face recognition technology

The definition provided by Article 29 Working Party enlightens three different tasks or functionalities for which FRT may be used³⁶⁴. First of all, FRT could be used for verification or authentication tasks (*one-to-one comparison*)³⁶⁵, meaning that the template is compared with *one* template in the database in order to generate a similarity score that reflects the likelihood that the person shown on the two samples is the same person. In verification or authentication tasks, the template must be matched against only one record, the one that corresponds to the declared identity³⁶⁶.

³⁶⁰ *Ibidem*.

³⁶¹ T. CHRISTAKIS, K. BANNELIER, C. CASTELLUCCIA, D. LE MÉTAYER, *Mapping The Use Of Facial Recognition In Public Spaces In Europe – Part. 2: Classification* (MAPFRE), Report of the AI-Regulation Chair, May 2022, 7.

³⁶² National Academies, *Facial Recognition Technology*, 33. This is especially evident in relation to face images collected under unconstrained circumstances, where lightening, facial expression and head position may be different.

³⁶³ This is not always the case in criminal investigations, where the FRT returns a list of photos of candidates similar to the probe image, which are presented to the police for a subsequent review.

³⁶⁴ FRT’s tasks can be found in FRA, *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, cited, 7-8. Differentiating between the diverse tasks is crucial for a proper legal and ethical debate on FRT. See CNIL, *Facial Recognition: For a Debate Living up to the Challenges*, cited. INTRONA, NISSENBAUM, *Facial Recognition Technology*, cited, 11.

³⁶⁵ Biometric verification is described as a «biometric product function that performs a one-to-one comparison». ISO/IEC 19792:2009, 4.3.12.

³⁶⁶ FRT for verification purposes can be used in passport control at the border. C. DUMBRAVA, *Artificial intelligence at EU borders: Overview of applications and key issues*, EPRS, European Parliament, July 2021.

Second, FRT could be used for identification purposes (*one-to-many comparison*)³⁶⁷. In this case, the template of a person's facial image is compared with *many* other templates stored in a database. In real-world identification problems, the operator does not necessarily know in advance whether the subject is in the reference database or not. This alternative gives rise to the difference between “closed-set identification problems” and “open-set identification problems”³⁶⁸. In addition, some authors distinguish different sub-categories of face identification applications, depending on the number of facial images acquired (and templates generated) that have to be compared with one or more templates stored in the reference database. More precisely, they distinguish between applications that only use one captured facial image (defined as “face search” – 1-M), applications that capture faces of all of the people in a specific venue (defined as “large scale face matching” – N-M) and applications that focus on a single reference face but which capture everyone in a specific venue (defined as “Targeted Face Tracking” - N-1). In the case of identification tasks, similarity scores are generated and those that are above the identified threshold are sorted from the highest to the lowest.

Third, FRT can be used for categorization purposes, which means to extract from a facial image information about the individual – such as gender, ethnic origin, age, facial expressions and emotional states³⁶⁹ – in order to classify individuals into groups according to their personal characteristics³⁷⁰. Although categorization does not imply any comparison or matching³⁷¹, it can be equally risky as it could point to characteristics of individuals that could, if linked to other information (e.g. location),

³⁶⁷ Biometric identification is described as a «*biometric system function that performs a one-to-many search to obtain a candidate list*». ISO/IEC 19792:2009, 4.3.9. Identification tasks include searching in a database for the identity of an unknown person or the monitoring of a person's movements.

³⁶⁸ In relation to the former, the human operator already knows that the person is included in the reference database. In relation to the latter, the human operator does not know in advance whether the person is in the reference database or not. See INTRONA, NISSENBAUM, *Facial Recognition Technology*, cited, 12. In the MAPFRE Project the authors distinguish between «*the basic features (“I”, “N” or “M”) of the inputs of the system, namely the “Captured Faces” (i.e. captured facial images) on the one hand and the “Reference Faces” (reference facial images) on the other hand*», specifying that «*“I” refers to a single face introduced as an input to the system, for instance the photo of a person taken at an eGate while passing through border control at the airport (example of “I” in “Captured Faces”); or the photo of this same person appearing in his or her biometric passport (example of “I” in “Reference Faces”). “N” refers to the faces of all persons present in a specific public space (for instance a station or a road) which are captured as an input by an FR system before processing them to check, for instance, if they match the faces in a database. “M” refers to the faces of a specific and pre-determined number of persons appearing in a reference database. Depending on the situation, “M” could refer to just a few persons (for instance Air France passengers boarding an airplane who are listed in the database of a FRT application such as “MONA” in France); hundreds of persons (for instance 1200 persons in a watchlist); or even to millions of persons (for instance those who appear in a general law enforcement database such as the TAJ in France)*». MAPFRE – Part. 2, cited, 11.

³⁶⁹ Among others, see J. BUOLAMWINI, V. ORDÓÑEZ, J. MORGENSTERN, E. LEARNED-MILLER, *Facial Recognition Technologies: A Primer*, Algorithmic Justice League, 2020, 2-6.

³⁷⁰ Inferring characteristics of individuals from facial images must be done carefully, remembering that FRT is based on correlation, not causation. X. WU, X. ZHANG, *Automated inference on criminality using face images*, 26 May 2017. arXiv:1611.04135v1.

³⁷¹ In MAPFRE – Part. 2 categorization is referred to as «*face analysis*». «*With face analysis, there is only one input, and there is no use of biometric templates, no face matching or identification. So, there is no “facial recognition” strictly speaking here*». *Ibidem*, 7.

allow identification of the individual or could lead to discrimination. For example, research shows that FRT from facial images can infer sexual orientation³⁷² or personality traits³⁷³. The distinction between functionalities is extremely important, as they pose different risks to human rights and freedoms³⁷⁴. As anticipated, besides the functionalities, the purposes for which FRT is used are also central³⁷⁵, as well as the operating conditions under which FRT is used. In addition to whether FRT is used in controlled or uncontrolled environments, and in public or non-public spaces, it is also important to consider whether it is deployed in real-time or retrospectively. Real-time face recognition (hereinafter also “live FRT”) refers to the instantaneous capture and analysis of facial features³⁷⁶, typically from live video streams in public spaces, with the aim of immediately matching individuals against a predefined watchlist. In contrast, non-real-time FRT involves a post-event analysis of stored images or video, usually for investigative purposes, and is generally subject to different legal standards. Live FRT is not only considered particularly intrusive – enabling continuous surveillance and real-time tracking of individuals – but also poses operational risks due to its demand for immediate intervention. The need for rapid decision-making by law enforcement officers increases the likelihood of acting on a false positive, thereby heightening the risk of detaining or arresting individuals who have been mistakenly identified by the system³⁷⁷. The differences highlighted above are particularly significant under the AI Act, which considers such distinctions when classifying the level of risk associated with a system.

³⁷² FRT is considered to be even better than humans in identifying individuals’ characteristics. Y. WANG, M. KOSINSKI, *Deep neural networks are more accurate than humans at detecting sexual orientation from facial images*, in *Journal of Personality and Social Psychology*, 114(2), 2018, 246.

³⁷³ A. KACHUR, E. OSIN, D. DAVYDOV, ET AL. *Assessing the big five personality traits using real-life static facial images*, in *Scientific Reports*, 10(1), 2020, Article 8487, 10.

³⁷⁴ As far as authentication/verification purposes are concerned, Annex III, 1(a) AI Act states that remote biometric identification systems are considered high-risk systems, but «*This shall not include AI systems intended to be used for biometric verification the sole purpose of which is to confirm that a specific natural person is the person he or she claims to be*». For a detailed analysis of FRT for authorisation purposes see T. CHRISTAKIS, K. BANNELIER, C. CASTELLUCCIA, D. LE MÉTAYER, *Mapping the Use of Facial Recognition in Public Spaces in Europe – Part 3: Facial Recognition for Authorisation Purposes*, Report of the AI-Regulation Chair, 2022.

³⁷⁵ Concerns are different if the FRT is used to identify a person suspected of committing a crime during criminal investigations, or if the FRT is used to identify a victim of a serious crime (e.g. kidnapping) or if the FRT is used to identify a face match on a large scale. «*To arrive at a satisfactory classification of the uses of FR in public spaces, it is therefore necessary, to consider both the technical elements (FPC functionalities and inputs) and the purpose for which the FR system is used (its “applications”) that will determine the outputs of the system and will greatly influence the risk assessment*». MAPFRE – Part. 2, cited, 9-10.

³⁷⁶ Information Commissioner’s Office, *Information Commissioner’s Opinion: The use of live facial recognition technology in public places*, June 2021.

³⁷⁷ In this terms A.A. GIKAY, *Regulating Use By Law Enforcement Authorities Of Live Facial Recognition Technology In Public Spaces: An Incremental Approach*, cited, 5.

2.2. The use of face recognition technology: where does Europe stand?

Although the European Union seems to be leaning toward a restrictive stance³⁷⁸, attempts to use FRT for law enforcement purposes are spreading across Member States³⁷⁹. At present, there is no systematic adoption of FRT by police departments; rather, its deployment is limited to individual use cases or pilot initiatives. In some cases, experimental projects have been conducted based on consensus, with all the delicate issues that arise in connection with this legal basis when public authorities are involved³⁸⁰.

Despite the growing debate on the use of FRT in Europe³⁸¹, very little is known about its actual implementation in the context of law enforcement³⁸². It seems that some countries, such as France and

³⁷⁸ I will analyse the provisions of the AI Act on biometric identification for law enforcement purposes in Chapter 3. At the EU level the European Data Protection Board and the European Data Protection Supervisor in their joint opinion on the AI Act proposal, called for a general ban of biometric identification in public spaces. See EDPB-EDPS, *Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, June 2021. In the U.S. context, some cities banned the use of FRT, with limited exceptions, also for law enforcement purposes. See City and County of San Francisco, *Board of Supervisors Approval of Surveillance Technology Policy*, San Francisco Administrative Code Section 19B.2(d), 2019. Available at: <https://www.sanfranciscopolice.org/your-sfpd/policies/19b-surveillance-technology-policies>.

³⁷⁹ Interesting reports on how facial recognition is being used and the related legal issues in the EU are published within the MAPFRE Research Project, a project established by the AI-Regulation Chair of the Multidisciplinary Institute in Artificial Intelligence (MIAI) at France's Université Grenoble Alpes. MAPFRE stands for «MAPping the use of Facial Recognition in public spaces in Europe». MAPFRE published three Reports which have shown a clear and detailed overview of the FRT's uses in Europe. T. CHRISTAKIS, K. BANNELIER, C. CASTELLUCCIA, D. LE MÉTAYER, *Mapping the Use of Facial Recognition in Public Spaces in Europe – Part 1: A Quest for Clarity: Unpicking the 'Catch-All' Term, - Part 2. Classification, - Part 3: Facial Recognition for Authorisation Purposes*, Reports of the AI-Regulation Chair, MIAI, May 2022. Many reports have been published by European Institutions or bodies. See T. MADIEGA, H. MILDEBRATH, *Regulating facial recognition in the EU. In-depth analysis*, September 2021. See also D. DUSHI, *The use of facial recognition technology in EU law enforcement: fundamental rights implications*, Global Campus of Human Rights, 2020.

³⁸⁰ The GDPR places certain limitations on the use of consent as a legal basis for the processing of personal data by public authorities and limitations are even stricter in the context of law enforcement. Recital n. 43 GDPR states that “(43) In order to ensure that consent is freely given, consent should not provide a valid legal ground for the processing of personal data in a specific case where there is a clear imbalance between the data subject and the controller, in particular where the controller is a public authority and it is therefore unlikely that consent was freely given in all the circumstances of that specific situation». Recital 35 LED states that “[...] The performance of the tasks of preventing, investigating, detecting or prosecuting criminal offences institutionally conferred by law to the competent authorities allows them to require or order natural persons to comply with requests made. In such a case, the consent of the data subject, as defined in Regulation (EU) 2016/679, should not provide a legal ground for processing personal data by competent authorities”. For these reasons, in connection with one pilot project, the CNIL opted for a blockade stating that consent «was not freely given and that the processing of biometric data in this case could not meet the requirements of necessity and proportionality». The pilot project specifically concerned the use of the FRT at the entrance of two high schools to allow only authorised students to enter the building and prevent theft or misuse of identity cards. The CNIL intervened, stressing the conditions for the correct implementation of such pilot project (CNIL, *Facial Recognition: For a debate living up to the challenges*, cited). After the position of the CNIL, opposing opinions emerged. Finally, the Administrative Court of Marseille confirmed the position of the CNIL. See Tribunal Administratif De Marseille, N° 1901249, Judgement February 27, 2020. T. CHRISTAKIS, *First Ever Decision of a French Court Applying GDPR to Facial Recognition*, AI-Regulation.com, February 27th, 2020.

³⁸¹ One of the most comprehensive work on FRT in Europe is C. LEQUESNE ROTH (Director), J. CAFFIN, M. KIMRI, M. KUBIAK, C. LACOUR, P. LEGROS, M. LEMOS, *La reconnaissance faciale dans l'espace – Une cartographie juridique européenne*, April 2020. See also *Biometric & Behavioural Mass Surveillance In EU Member States*, Report for the Greens/EFA in the European Parliament, October 2021.

³⁸² This is recognised by JANSEN, *Data-driven policing: Negotiating the legitimacy of the police*, cited, 114. An interesting study on FRT in Europe is L. MONTAG ET AL., *The Rise and rise of biometrics mass surveillance in the EU, A legal analysis*

Germany, are more inclined to use FRT than others. In Germany, for example, the Berlin Police conducted a trial of live FRT at the train station Berlin-Südkreuz from August 2017 to July 2018. The purpose of the test was to assess the technical capabilities of FRT. The watchlist consisted exclusively of volunteers, and the section of the train station observed by the system was clearly defined³⁸³. Once the test was completed, the police published a report on the test results in September 2018³⁸⁴. Similarly, the Nice's police tested live FRT during the 2019 Nice carnival to assess its technical capabilities³⁸⁵. The watchlist comprised only volunteers, and – as in the German test – they could choose whether or not to enter the area where the system was deployed³⁸⁶. In addition to such pilot projects, during the 2017 G20 Summit in Hamburg, the police actually used FRT for detection and investigation of crimes. In Italy, the debate related to the use of FRT deserves special reference, as there has been an unclear trend over the years. In 2017, the Italian Ministry of Interior – Department of Public Security started using SARI (Automatic System of Image Recognition). As stated in the technical specifications³⁸⁷, the software was developed to operate in two different scenarios: SARI Enterprise and SARI Real-Time. SARI Enterprise is aimed at identifying a person depicted in an image by analysing a large database, the Automated Fingerprint Identification System (AFIS). SARI Enterprise can thus be considered a face search (1-M) application. By contrast, the objective of SARI Real-time is to analyse in real time faces of subjects captured by cameras installed in a circumscribed area and compare them with a predefined database (called watch-list). SARI Real-time therefore can be considered a Large Scale Face Matching (N-M) application³⁸⁸.

Apart from actual Member States, the United Kingdom is probably the European country where police forces have developed and deployed FRT the most³⁸⁹. The Metropolitan Police Service and South

of biometrics mass surveillance practices in Germany, the Netherlands, and Poland, EDRI – European Digital Rights, 2021.

³⁸³ A detailed description of this pilot project can be found in A.V. EIREINER, *Imminent dystopia? Media coverage of algorithmic surveillance at Berlin-Südkreuz*, in *Internet Policy Review*, 9(1), 2020. Available at: <https://policyreview.info/articles/analysis/imminent-dystopia-media-coverage-algorithmic-surveillance-berlin-sudkreuz>.

³⁸⁴ Bundespolizeipräsidium Potsdam, *Teilprojekt 1 – Biometrische Gesichtserkennung*, 18 September 2018. Available at: https://www.bundespolizei.de/Web/DE/04Aktuelles/01Meldungen/2018/10/181011_abschlussbericht_gesichtserkennung_down.pdf?__blob=publicationFile.

³⁸⁵ FRA, *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, cited, 12.

³⁸⁶ For a detailed description see *Biometric & Behavioural Mass Surveillance In EU Member States*, cited, 83-88.

³⁸⁷ Ministero dell'Interno – Dipartimento di Pubblica Sicurezza, *Capitolato tecnico, Procedura volta alla fornitura della soluzione integrata per il sistema automatico di riconoscimento immagini S.A.R.I.* Available at: <https://www.poliziadistato.it/statics/06/20160627-ct-sari--4-.pdf>.

³⁸⁸ While SARI Enterprise has been in use since September 2018, SARI Real-time is not yet in use.

³⁸⁹ S. DEL GATTO, *Riconoscimento facciale. Secondo uno studio la polizia del Regno Unito non rispetta gli standard minimi etici e legali*, published in *Osservatorio sullo Stato Digitale*, IRPA, 22 February 2023. In the United Kingdom, a research group of the Minderoo Centre for Technology and Democracy by the University of Cambridge has elaborated an audit mechanism to evaluate the compliance with the law and national guidance of police use of FRT, with a specific focus on issues of privacy, equality and freedom of expression and assembly. Researchers elaborated minimum ethical and legal standards in order to assess, through a series of questions, the quality of FRT governance by English police forces and then applied them to three case studies. The first one is the aforementioned *R (Bridges) v. Chief Constable of South Wales Police*,

Wales Police have conducted trial deployments of live FRT³⁹⁰, and have publicly released the results of these trials³⁹¹. More specifically, the Metropolitan Police Service conducted multiple FRT tests from 2016 to 2019. In the independent report authored by Professor Pete Fussey and Dr. Daragh Murray³⁹² the authors identified ten pilot projects³⁹³. They highlighted significant concerns in relation to the Metropolitan Police Service's uses of live FRT³⁹⁴. Similarly, the South Wales Police used live FRT at a variety of large public events, such as football matches (e.g. at the UEFA Champions League Final in June 2017) or concerts³⁹⁵.

3. Considerations on Automated Video Surveillance and Face Recognition Technology

Similar to predictive policing, automated video surveillance and face recognition technologies give rise to a number of concerns.

3.1. Data and data protection concerns

Currently, huge amounts of personal data are collected, stored and processed through AI-based surveillance devices, underpinning serious challenges to the right to protection of personal data

the second concerns the use of live facial recognition by the Metropolitan Police Service (MPS) from August 2016 to February 2019, and the last one relates to the use of FRT by South Wales Police from December 2021 to March 2022. E. RADIYA-DIXIT, *A Sociotechnical Audit: Assessing Police Use Of Facial Recognition*, Minderoo Centre for Technology and Democracy, University of Cambridge, October 2022.

³⁹⁰ B. DAVIES, M. INNES, A. DAWSON, *An evaluation of South Wales Police's use of Automated Facial Recognition*, Police Science Institute, Crime and Security Research Institute, Cardiff University, September 2018. Practices for the use of live FRT by the Metropolitan Police Service and the South Wales Police are established by the Standard Operating Procedures. See Metropolitan Police, *Standard Operating Procedure for the Overt Deployment of Live Facial Recognition*, November 2022. Available at: <https://www.met.police.uk/SysSiteAssets/media/downloads/force-content/met/advice/lfr/policy-documents/lfr-sop.pdf>.

³⁹¹ MPS LFR Deployments 2020 – 2022. Available at: <https://www.met.police.uk/SysSiteAssets/media/downloads/force-content/met/advice/lfr/new/lfr-deployment-grid-2020-2022.pdf>. T. MANSFIELD, *Facial Recognition Technology in Law Enforcement Equitability Study Test Strategy*, National Physical Laboratory, April 2022.

³⁹² Professor Pete Fussey and Dr. Daragh Murray are members of the ESRC Human Rights, Big Data & Technology Project, at the University of Essex Human Rights Centre.

³⁹³ P. FUSSEY, D. MURRAY, *Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology*, Human Rights Centre, University of Essex, 15. Metropolitan Police Service, *Live Facial Recognition: Legal Mandate*, 29 November 2022. Accessible at: <https://www.met.police.uk/SysSiteAssets/media/downloads/force-content/met/advice/lfr/policy-documents/lfr-legal-mandate.pdf>.

³⁹⁴ Among the concerns, the most relevant ones are: the research process adopted by Metropolitan Police Service to trial live FRT; the absence of an explicit legal basis for the use of live FRT; the absence of an analysis on the necessity requirement as required by human rights law; the procedures and practices of live FRT in operational settings. *Ibidem*, 15.

³⁹⁵ South Wales Police, *New Facial Recognition Mobile App to Identify Vulnerable, Missing and Wanted Individuals*, 7 December 2021. Available at: [New facial recognition mobile app to identify vulnerable, missing and wanted individuals | South Wales Police \(south-wales.police.uk\)](https://www.south-wales.police.uk/news/new-facial-recognition-mobile-app-to-identify-vulnerable-missing-and-wanted-individuals).

enshrined in the Charter on Fundamental Rights of the European Union (hereinafter “CFREU”) and in the European Convention on Human Rights (hereinafter “ECHR”). From the perspective of data protection legislation, the processing of personal data can be identified whenever a natural person is filmed by video surveillance cameras³⁹⁶. Therefore, AVS systems generally process personal data, since a person’s face can be captured even when an AVS is used for the sole purpose of detecting the presence of an object in a video stream, and provisions enshrined in data protection legislation apply. One of the most critical aspects is that data subjects are not always aware that their data is being processed by AI-based devices.

In many cases DPAs have intervened in relation to AI-based software due to the absence of a specific legal basis for their use. In the specific context of law enforcement, in July 2018, the Italian DPA found that the Ministry of the Interior’s SARI Enterprise system complied with national and European data protection requirements³⁹⁷. The rationale was that its functionalities were limited to automating specific tasks: users could search a database of photo-identified individuals by uploading an image, which the system automatically analysed to produce a list of similar mugshots³⁹⁸. The Italian DPA clarified that SARI Enterprise does not constitute a new processing operation but merely a new method for processing existing personal data, and therefore the existing legal basis was considered sufficient. By contrast, in 2021 the DPA found that SARI Real-Time was incompatible with national and European law, as there was no adequate or specific legal basis to authorise the processing of biometric data through a system operating as a Large-Scale Face Matching (N–M) application³⁹⁹. In particular, through a network of cameras installed within a defined geographical area, the system can analyse individuals’ faces in real time and compare them against a predefined database (the “watch-list”), which may contain up to 10,000 facial images. When the face recognition algorithm identifies a match between a face on the watch-list and a face captured by one of the cameras, it generates an alert to notify law enforcement officers. Moreover, the system can record the footage captured by cameras, thereby also functioning as a video-surveillance tool.

³⁹⁶ Garante per la Protezione dei Dati Personali, Video Surveillance Decision dated 8 April 2010.

³⁹⁷ Garante per la Protezione dei Dati Personali, Decision No. 440/2018, 26 July 2018.

³⁹⁸ “*Il trattamento in argomento costituisce [...] un mero ausilio all’agire umano, avente lo scopo di velocizzare l’identificazione, da parte dell’operatore di polizia, di un soggetto ricercato della cui immagine facciale si disponga, ferma restando l’esigenza dell’intervento dell’operatore per verificare l’attendibilità dei risultati prodotti dal sistema automatizzato*”. *Ibidem*.

³⁹⁹ Garante per la Protezione dei Dati Personali, Decision No. 127/2021, 25 March 2021. See A. MASCOLO, *Riconoscimento facciale “in tempo reale”: il Garante per la privacy bocchia S.A.R.I. Real Time*, published in *Osservatorio sullo Stato Digitale*, IRPA, 22 July 2021.

Data protection concerns are even stronger when FRT or other biometric technologies are involved, as the processing of biometric data comes into play⁴⁰⁰. The five-step process described above elucidates that facial images cannot automatically be considered “biometric data” under data protection legislation. A face captured by a traditional surveillance camera or in a standard photograph does not constitute biometric data. It becomes biometric data only when “*specific technical processing*” is applied to extract identifying characteristics from the image to uniquely identify the individual⁴⁰¹. Data protection concerns relate to various stages of handling facial images in FRT systems. These include the collection, processing, and retention of facial images used for training ML-models, as well as those used as probe images and reference images⁴⁰².

The definition of biometric data emphasises the importance of specific technical processing, which distinguishes the mere processing of personal data (e.g., a facial image) from the processing of biometric data (e.g., the use of FRT to identify an individual from that image). However, there remain some uncertainties on the scope and meaning of this concept⁴⁰³. DPAs have thus issued guidance to bridge this gap⁴⁰⁴, specifying that when a facial image is prepared for matching (biometric feature extraction or biometric template generation), this constitutes biometric data processing⁴⁰⁵. Although

⁴⁰⁰ Article 4(14) GDPR, Article 3(13) LED and Article 3(18) Regulation 2018/1725 define biometric data as «*personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data*». For a discussion on the difference between the term «biometrics» and «biometric data» - the former as a technical notion, the latter as a legal and technical one - see A.C. JASSERAND, *Avoiding terminological confusion between the notions of 'biometrics' and 'biometric data': an investigation into the meanings of the terms from a European data protection and a scientific perspective*, in *International Data Privacy Law*, 6(1), 2016, 63–76.

⁴⁰¹ As articulated in Recital 51 of the GDPR, the processing of photographs «*should not systematically be considered to be processing of special categories of personal data as they are covered by the definition of biometric data only when processed through a specific technical means allowing the unique identification or authentication of a natural person*». See European Data Protection Board, *Guidelines 3/2019 on Processing of personal data through video devices*, 2020. See G. GONZÁLEZ FUSTER, *Artificial Intelligence and Law Enforcement Impact on Fundamental Rights*, July 2020, 24. FRA, *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, cited, 5.

⁴⁰² Collection refers to how facial images are initially obtained, whether from surveillance cameras, photo uploads, social media, or public databases. Processing involves the use of collected images for facial recognition tasks such as detection, identification, and verification. It includes feature extraction, algorithm training, and comparisons made during recognition attempts. Retention pertains to how long and where these images are stored.

⁴⁰³ For technical discussion see B.O. OMOYIOLA, *Overview of Biometric and Facial Recognition Techniques*, in *Journal of Computer Engineering*, 20(4), 2018, 1-5.

⁴⁰⁴ «*The term “specific technical processing” describes a processing operation – or set of operations – that can be applied to a person’s physical, physiological or behavioural characteristics, which makes it possible to uniquely identify them*». The term «specific technical processing» can also «*refer to the main stages involved in biometric recognition systems. For example: biometric feature extraction, where the information in the biometric sample is extracted and transformed by an algorithm into a biometric feature; and biometric template generation, where a biometric feature is stored as a biometric Template*». ICO, *Biometric recognition*, Available at: <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/biometric-data-guidance-biometric-recognition/biometric-recognition/#specific>.

⁴⁰⁵ The EDPB explicitly states that «*In order for it to be considered as processing of special categories of personal data (Article 9) it requires that biometric data is processed “for the purpose of uniquely identifying a natural person”. To sum up, in light of Article 4.14 and 9, three criteria must be considered: - Nature of data: data relating to physical, physiological or behavioural characteristics of a natural person, - Means and way of processing: data “resulting from a specific technical processing”, - Purpose of processing: data must be used for the purpose of uniquely identifying a natural person*». EDPB, *Guidelines 3/2019 on processing of personal data through video devices*, cited.

doubts remain on the boundaries of the notion of “biometric data” under data protection legislation, the distinction between personal data and biometric data is crucial, as it carries important implications for the regulatory frameworks and legal obligations governing such processing: in the case of biometric data processing, stricter protections and obligations are set for the collection, storage and processing of such data⁴⁰⁶.

On this specific topic, the Italian DPA intervened on an initiative by the Italian municipality of Como, which deployed a facial-recognition-enabled video-surveillance system in the green area known as Parco Tokamakhi⁴⁰⁷. According to the municipality, the system aimed to assist the Judiciary Police unit in identifying individuals under investigation or reported missing, automatically detect suspicious behaviour (e.g., loitering) or potentially dangerous situations, and identify thefts. The DPA concluded that this form of biometric data processing is permissible solely where an appropriate and specific legal basis exists, which was not in place at the time. A few months later, in 2021, the Italian Government adopted an urgent measure – *i.e.* Decree Law No. 139/2021, subsequently converted into Law No. 205/2021 – that introduced, among other elements, a moratorium on the use of face recognition in Italy. The measure suspended the deployment and use of video-surveillance systems incorporating such technology until the adoption of specific legislation and, in any case, until December 2023 and later extended to December 2025⁴⁰⁸. However, it was later clarified that the moratorium did not extend to processing activities conducted by competent authorities for the prevention and suppression of crime or the execution of criminal sanctions, on the condition that a favourable opinion is first obtained from the DPA⁴⁰⁹. Nevertheless, lacking a legal basis, the DPA blocked two further projects submitted by

⁴⁰⁶ The reference goes to Article 9 GDPR and Article 10 LED. I will deal with the data protection regulatory framework in Chapter 3. Some authors highlight that the reference to the specific technical processing was added in a final phase of the GDPR, as a result of a political agreement in the Council, when the General Approach was adopted in June 2015. See JASSERAND-BREEMAN, *Reprocessing of biometric data for law enforcement purposes: Individuals’ safeguards caught at the Interface between GDPR and the “Police” directive?*, cited, 2019, 46-54. The reason for this addition, for another author, is to exclude the processing of databases which contain facial images, generally for the purpose of law enforcement, from the special protections accorded to the processing of biometric data. E. J. KINDT, *Having yes, using no? About the new legal regime for biometric data*, in *Computer Law and Security Review*, 34(3), 2018, 12.

⁴⁰⁷ Garante per la protezione dei dati personali, Decision No. 54/2020, February 26, 2020.

⁴⁰⁸ Article 9(9) Law No. 205/2021; Article 8-ter Decree Law No. 51/2023. Following the entry into force of the AI Act, the moratorium must be coordinated with the Act’s provisions specifically governing the use of biometric identification.

⁴⁰⁹ Article 9(12) Law No. 205/2021. The provision, however, was far from clear. The government’s original intention was to prohibit the use of facial recognition technology in the absence of an appropriate and specific legal framework. Moreover, the Italian Data Protection Authority had previously stated – in Decision No. 127/2021 – that no favourable opinions could be issued until such a framework was formally adopted. M. R. CARBONE, *SARI, il riconoscimento facciale nella pubblica sicurezza: servono regole e trasparenza* published in *Agenda Digitale EU*, 23 March 2021. Available at: <https://www.agendadigitale.eu/sicurezza/privacy/sari-vantaggi-e-rischi-del-riconoscimento-facciale-nella-pubblica-sicurezza/>.

the municipalities of Lecce and Arezzo, both of which were preparing to implement video-surveillance systems integrated with FRT⁴¹⁰.

Similar considerations have been made by DPAs around Europe. With respect to the use of FRT for crime detection and investigation during the 2017 G20 Summit in Hamburg, the Hamburg DPA found that the video-surveillance measures and the ensuing extraction and retention of biometric data were not supported by an adequate legal basis⁴¹¹. Although a first-instance court overturned the initial order requiring deletion of the police database containing biometric templates, the Hamburg DPA maintained on appeal that the lack of a clear and specific legal basis amounted to a breach of Article 8(2) CFREU, Article 4(1)(a) LED, and the relevant national provisions⁴¹². Similarly, the use of Automated Facial Recognition (AFR) by South Wales Police led to the first legal challenge concerning the use of face recognition technology for law enforcement purposes in the United Kingdom, culminating in the landmark case *R. (Bridges) v Chief Constable of South Wales Police*⁴¹³. The underlying question was whether there was a sufficient legal framework for the use of AFR by the police. In this case, the Divisional Court in Cardiff initially ruled that the police's deployment of AFR in public spaces was lawful, finding that the existing legal framework was sufficient and that its use remained within legal bounds⁴¹⁴. However, this ruling was later overturned by the Court of Appeal⁴¹⁵,

⁴¹⁰ Garante per la Protezione dei dati personali, *Videosorveglianza: stop del Garante privacy a riconoscimento facciale e occhiali smart. L'Autorità apre istruttorie nei confronti di due Comuni*, 14 November 2022. See M. GIUSTI, *Stop del Garante privacy ai progetti di videosorveglianza smart di Lecce e di Arezzo*, published in *Osservatorio sullo Stato Digitale*, IRPA, 8 February 2023.

⁴¹¹ Hamburg DPA, *Datenschutzrechtliche Prüfung des Einsatzes einer Gesichtserkennungssoftware zur Aufklärung von Straftaten im Zusammenhang mit dem G20-Gipfel durch die Polizei Hamburg*, August 31, 2018.

⁴¹² T. MADIEGA, H. MILDEBRAT, *Regulating facial recognition in the EU*, cited, 12. R. VENEMA, *How to Govern Visibility?: Legitimizations and Contestations of Visual Data Practices after the 2017 G20 Summit in Hamburg*, in *Surveillance & Society*, 2020, 18(4), 522-539. M. BRÖCKLING, *Gerichtsurteil zu Gesichtserkennung: Datenschützer scheitert an Löschung biometrischer G20-Datenbank*, published in *Netzpolitik.org*, 24 October 2019. Available at: <https://netzpolitik.org/2019/datenschuetzer-scheitert-an-loeschung-biometrischer-g20-datenbank/>.

⁴¹³ UK, High Court of Justice (Queens' Bench Division – Divisional Court Cardiff), *The Queen (OTAO) Bridges and Chief Constable of South Wales Police and others*, [2019] EWCH 2341 (Admin), *R (Bridges) v CCSWP and SSHD*, September 4, 2019. The claimant complained of the violation of the Data Protection Act, of the ECHR and of the Equality Act. Specifically, the claimant submits that the processing of biometric data by video-cameras located in public spaces involved people who were not suspected of having committed (or being about to commit) criminal offences, thus violating their right to respect for private and family life and the protection of personal data. After going over all the applicable legislation, the Court dismissed the claims, stating that the current UK national legal regime was adequate to ensure appropriate and non-arbitrary use of FRT and that the use of FRT (AFR Locate) by the police was consistent with the applicable legal framework. S. DEL GATTO, *Quali regole per le nuove tecnologie di riconoscimento facciale? La Corte di giustizia di Cardiff si pronuncia per la legittimità dell'uso di tecniche di Automated Facial Recognition da parte della Polizia del Galles*, published in *Osservatorio sullo Stato Digitale*, IRPA, 21 April 2020.

⁴¹⁴ "The legal framework that we have summarised does provide a level of certainty and foreseeability that is sufficient to satisfy the tenets of Article 8(2). It provides clear legal standards to which SWP will be held". UK, High Court of Justice, cited, paragraph 96.

⁴¹⁵ Court of Appeal, Civil Division, *R (Bridges) v. Chief Constable of South Wales Police*, Judgement, case C1/2019/2670, August 11, 2020. It is sufficient to recall that in August 2020, the Court of Appeal ruled that the use of live facial recognition by South Wales Police was unlawful. The Court held that there were «fundamental deficiencies [...] in the legal framework currently in place» and, specifically, «too much discretion is currently left to individual police officers. It is not clear who can be placed on the watchlist nor is it clear that there are any criteria for determining where AFR can be deployed

which held that the use of AFR Locate failed to meet key legal requirements due to the inadequacy of the governing legal framework⁴¹⁶. At present, there is no legal basis authorising the general use of FRT at the national level. This absence is problematic not only under data protection law, which requires a lawful basis for the processing of biometric data, but also in terms of justifying interferences with fundamental rights⁴¹⁷. In any event, the use of FRT for law enforcement purposes must now comply with the AI Act, which establishes specific rules governing such use. In particular, it requires Member States to set out in their national law detailed provisions authorising the use of “real-time” remote biometric identification systems in publicly accessible spaces for law enforcement purposes, strictly within the limits and subject to the conditions laid down in Article 5(1), first subparagraph, point (h), and paragraphs 2 and 3.

In addition, data protection issues arise when law enforcement authorities rely on datasets collected for different purposes to train or operate their systems. This includes, for instance, data gathered during criminal investigations or data obtained by other public authorities for administrative purposes, such as border control management, or data held by private parties. Such practices often occur without the knowledge or consent of individuals concerned, raising serious questions about compliance with data protection principles, especially regarding purpose limitation⁴¹⁸. Similar practices have been evident in the context of Member State authorities responsible for internal security gaining access to EU centralised databases, such as the Visa Information System (VIS) and European Dactyloscope (EURODAC), which were originally created for administrative purposes. However, law enforcement authorities have increasingly sought access to them for policing and security-related objectives. In response to this trend, the European Data Protection Supervisor (EDPS) issued a warning as early as 2005, highlighting the risks associated with the systematic access of law enforcement authorities to these large-scale information systems for purposes other than those for which they were created. The EDPS stressed that such access could undermine core data protection principles, particularly the principle of purpose limitation, which requires that personal data is used only for the purposes for

identified key gaps in the police force’s compliance with human rights, equality, and data protection laws». *Ivi*, paragraph 91.

⁴¹⁶ Indeed, in August 2020, the Court of Appeal ruled that the use of live facial recognition by the South Wales Police was unlawful. In particular, the Court held that there were «*fundamental deficiencies [...] in the legal framework currently in place*» and, specifically, it was not clear “*who can be placed on the watchlist nor is it clear that there are any criteria for determining where AFR can be deployed identified key gaps in the police force’s compliance with human rights, equality, and data protection laws*”. *Ibidem*, paragraph 91.

⁴¹⁷ Interferences with fundamental rights are permissible only if they comply with the requirements set out in the CFREU and the ECHR. According to the three-part test developed by the ECtHR, any such interference must pursue a legitimate aim, be prescribed by law, and be necessary in a democratic society, including meeting the requirements of necessity and proportionality.

⁴¹⁸ MADIEGA, MILDEBRAT, *Regulating facial recognition in the EU*, cited, 8.

which it was originally collected. A further concern lies in the nature of individuals included in these databases: they are, in most cases, not suspected of any criminal activity⁴¹⁹.

3.2. Errors, accuracy concerns and bias

As previously noted, AVS and FRT rely on computer vision, a subfield of AI based on machine learning and deep learning, for extracting features from facial images and comparing extracted features with those stored in a database. As discussed in Chapter 1, a key characteristic of these non-deterministic algorithms is that they operate on a probabilistic (or stochastic) basis, meaning that their outputs are not fixed but depend on statistical patterns learned from data⁴²⁰. In order for FRT to perform the recognition task, a model must be built. The learning process specifically requires a training dataset, consisting of facial images, against which the chosen face recognition algorithms are applied. The creation of the training dataset is therefore crucial, as the performance and the accuracy of FRT highly depends on its quality⁴²¹. However, the importance of high-quality data is not limited to training datasets, rather it is crucial also for both the images stored in the reference database⁴²² and the probe images acquired for comparison, otherwise the performance of the FRT might be compromised⁴²³. Standards for minimum image quality (training, probe and reference images) therefore should be established.

⁴¹⁹ European Data Protection Supervisor, *Opinion of the European Data Protection Supervisor on the Proposal for a Council Decision concerning access for consultation of the Visa Information System (VIS) by the authorities of Member States responsible for internal security and by Europol for the purposes of the prevention, detection and investigation of terrorist offences and of other serious criminal offences (COM (2005) 600 final)*, 2006/C 97/03. European Data Protection Supervisor, *Opinion of the European Data Protection Supervisor on the amended proposal for a Regulation of the European Parliament and of the Council concerning the establishment of 'Eurodac' for the comparison of fingerprints for the effective application of Regulation (EC) No (.../...) (establishing the criteria and mechanisms for determining the Member State responsible for examining an application for international protection lodged in one of the Member States by a third-country national or a stateless person), and on the proposal for a Council Decision on requesting comparisons with Eurodac data by Member States' law enforcement authorities and Europol for law enforcement purposes*, 2010/C 92/01. The importance of individuals whose data are used was also underlined by the German Constitutional Court in its assessment of hessenDATA. "The systems also contain data on the persons who file complaints or provide information, as well as data on witnesses, persons involved in accidents, and other persons who are not suspects or accused persons within the meaning of criminal procedural law and who are not responsible for causing danger within the meaning of police law ([...]). In Hesse, this data is automatically integrated into the analysis platform ([...]). Bundersverfassungsgericht, Judgement 16 February 2023, cited, paragraph 134.

⁴²⁰ Consultative committee Convention 108, *Facial Recognition: Current Situation and Challenges*, cited, 3.

⁴²¹ FRA, *Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights*, Publications Office of the European Union, 2019, 2. «[H]igh-quality data are essential for the development of Artificial Intelligence». See European Council, *Council's conclusions from 28 June 2018*, EUCO 9/18, Par. 20. 5 C. GARVIE, *Garbage In, Garbage Out: Face Recognition on Flawed Data*, published in *Georgetown Law*, Center on Privacy and Technology, 16 May 2019. Available at: <https://www.law.georgetown.edu/privacy-technology-center/publications/garbage-in-garbage-out-face-recognition-on-flawed-data/>.

⁴²² Which can be a general database, a specific database or a specific watchlist.

⁴²³ INTRONA, NISSENBAUM, *Facial Recognition Technology A Survey of Policy and Implementation Issues*, cited, 17-19. National Academies, *Facial Recognition Technology*, cited, 1. MOBILIO, *Tecnologie di riconoscimento facciale*, cited, 12.

Capturing a high-quality image (a “standard image”⁴²⁴) is far from simple. It would require capturing the face image in controlled circumstances with the knowledge and cooperation of the subject⁴²⁵. However, such cooperative capture is unlikely the case in the context of law enforcement: in most cases facial images are captured without knowledge of the subject by CCTV cameras in outdoor scenes with unpredictable lights, shadows, orientations and complex backgrounds⁴²⁶, eventually resulting in lower-quality images. In different circumstances, such as in sport arenas or in airports, the situation might be different, and images can be of higher quality.

Although accuracy has increased significantly in recent years and advances in AI now make it possible to address challenges arising from variations in facial expressions and positions, aging or changes in facial appearance, lighting and low image resolution, sensitivity to these factors remains an issue⁴²⁷. First and foremost, FRT may either correctly identify a match – or produce a high similarity score – between the probe image and one or more images in the database, or correctly identify the absence of a match, resulting in a low similarity score. These outcomes correspond, respectively, to true positives (TP) and true negatives (TN). Conversely, the system may incorrectly identify a match (or produce a high similarity score), where there is none, or fail to recognise an existing match, resulting in false positives (FP) and false negatives (FN), respectively. More specifically, false negatives reflect system’s failures in associating two images of the same person, whereas false positives involve the erroneous association of images belonging to two different individuals⁴²⁸. In relation to one-to-many comparisons, when dealing with open-set identification problems, the absence of a match does not necessarily mean that an error has occurred, but rather that the subject in question might not be present in the database. In this case, therefore, it is crucial to distinguish between a true negative (i.e. the person concerned is not in the database) and a false negative (i.e. the person is in the database, but the FRT has made a mistake), as subsequent actions must be taken accordingly⁴²⁹.

⁴²⁴ NIST, *Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information*, NIST Special Publication 500-290, January 2011.

⁴²⁵ For a discussion on illumination issues see Y.A. MOSES, S. ULLMAN, *Face Recognition: The Problem of Compensating for Changes in Illumination Direction*, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 1997, 721-732.

⁴²⁶ In some cases, acquiring the probe image from a video might be better for pose and illumination variations, since the subject is captured multiple times with potentially different poses and illumination. See L. Introna, H. Nissenbaum, 19. Other strategies described by the authors to overcome the problems of pose and illuminations are the three-dimensional input and the infra-red input. Identification in uncontrolled environments might determine a dramatic loss in performance, as stated in J. GALBALLY, P. FERRARA, R. HARAKSIM, A. PSYLLOS, L. BESLAY, *Study on Face Identification Technology for its Implementation in the Schengen Information System*, JRC-34751, July 2019, 67.

⁴²⁷ The evolution of FRT makes error rates and variations across demographic groups increasingly smaller. National Academies, *Facial Recognition Technology*, cited, 32. An interesting question is «How accurate are humans?». See *Ibidem*, 63.

⁴²⁸ Other types of errors can be identified. In relation to the image acquisition phase, failure to acquire the image is a possible error, while failure to extract the feature is an error in the feature extraction phase.

⁴²⁹ INTRONA, NISSENBAUM, *Facial Recognition Technology A Survey of Policy and Implementation Issues*, cited, 13.

False negatives (i.e. failure to associate two images) generally derive from low-quality images. Pictures taken under controlled circumstances, e.g. mugshots or standard facial images in travel documents⁴³⁰, are generally considered of high quality and lead to better results⁴³¹. Conversely, low lighting, blurry, low-resolution cameras and motion may be factors that lead to low-quality images that consequently determine an increase in false negatives error rates. By contrast, false positives (i.e. incorrect match of templates that relate to two different individuals) are generally due to similarity between two images, in cases where anatomic similarities can be identified, such as in relation to identical twins or relatives⁴³².

There is always a trade-off between false negatives and false positives error rates⁴³³, which depends on the specified threshold⁴³⁴. The definition of the threshold reflects a value judgment that takes into account several variables, including the specific context in which the system operates, its intended purpose, and the potential consequences of false identifications⁴³⁵. Although recent technological advancements have significantly improved the performance of FRT systems, especially in challenging conditions such as low light or partial obstructions⁴³⁶, the trade-off between false positives and false negatives remains unavoidable⁴³⁷.

Both types of errors carry distinct risks. On the one hand, false positives can lead to wrongful identification, resulting in unjustified stops, searches, interrogations, or even arrests. These outcomes are particularly problematic in one-to-many matching scenarios, where a single individual is compared against a large database. On the other hand, false negatives, while detrimental to public security by allowing individuals to evade identification, may benefit the person by preventing detection. Defining an acceptable error rate thus requires careful consideration of both the operational context and the

⁴³⁰ International Civil Aviation Organization (ICAO), *Technical Report. Portrait Quality, Reference Facial Images for MRTD*, April 2018; International Standard Organization (ISO), International Electrotechnical Commission (IEC), *Standard for Best Practices for Face Images*, ISO/IEC 19794-5.

⁴³¹ Face images are considered of high-quality if they respect the standard ISO/IEC 39794-5:2019, *Information technology — Extensible biometric data interchange formats. Part 5: Face image data*. Available at: <https://www.iso.org/standard/72156.html>.

⁴³² In one-to-many identification, when the database is huge, it is easier that false positives may occur. National Academies, *Facial Recognition Technology*, cited, 52.

⁴³³ National Academies, *Facial Recognition Technology*, cited, 37-38.

⁴³⁴ An exception to this statement is the investigative use of FRT systems, where the threshold is generally set to zero. National Academies, *Facial Recognition Technology*, cited, 49.

⁴³⁵ Depending on how the threshold is defined, the system will provide different results. MOBILIO, *Tecnologie di riconoscimento facciale*, cited, 36.

⁴³⁶ The quality of cameras is much improved and, as a result, the quality of face images is now higher. Of course, the differences between the different types of cameras remain relevant. For example, differences can be seen between cameras which have a build-in capability in detecting faces and those who have not. In the former case, the camera generally focuses on the face, leading to higher quality face images. National Academies, *Facial Recognition Technology*, cited, 42-43.

⁴³⁷ The existence of a trade-off between false positives and false negatives is generally acknowledged. See FRA, *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, cited, 9; E. LEARNED-MILLER, V. ORDÓÑEZ, J. MORGENSTERN, J. BUOLAMWINI, *Facial Recognition Technologies in the Wild: A Primer*, cited, 11; MOBILIO, *Tecnologie di riconoscimento facciale*, cited, 36.

system's objectives⁴³⁸. These considerations are equally relevant to applications of AVS, other than those based on FRT.

Poor performance of FRT (high FP or FN rates) represents a fundamental challenge that significantly affects its reliability. In addition, variations of FP or FN rates across demographic groups currently represent a central issue⁴³⁹. FRT systems, like other machine learning- and deep learning-based systems, tend to be less accurate in identifying individuals with darker skin tones or distinct facial features, mainly due to the fact that they display different characteristics from those represented in datasets predominantly used to train face recognition algorithms, *i.e.* lighter-skinned male individuals⁴⁴⁰. Moreover, darker toned faces reduce feature contrast and FRT systems might fail to correctly identify the individual. In addition, poor-quality images might increase variations of false positives and false negatives rates across demographics⁴⁴¹.

Gender and race biases are the most frequently documented⁴⁴². In its comprehensive 2019 Face Recognition Vendor Test, the NIST conducted an in-depth analysis to determine whether and to what extent face recognition algorithms perform differently across various demographic groups⁴⁴³. The study found that false positive rates were significantly higher for individuals from West and East African background, as well as East Asian ones, while the lowest rates were observed among individuals of Eastern European descent⁴⁴⁴. These disparities were not marginal; conversely, the differences were substantial, indicating that certain racial or ethnic groups are at greater risk of being misidentified by FRT systems⁴⁴⁵. In addition, the NIST's report highlighted that false positives were more prevalent among women than men, a pattern that held consistently across multiple algorithms

⁴³⁸ In this sense, MOBILIO, *Tecnologie di riconoscimento facciale*, cited, 37. Indeed, AlgorithmWatch correctly stresses that «Even if face recognition can match a face with 99% accuracy, the sheer amount of faces available in police databases makes false positives inevitable. (The 1% error rate means that, if 10,000 people who are not wanted by the police undergo face recognition, 100 will be flagged as wanted)». N. KAYSER-BRIL, *At least 11 police forces use face recognition in the EU, AlgorithmWatch reveals*, Algorithm Watch, 11 December 2019. Accessible at: <https://algorithmwatch.org/en/face-recognition-police-europe/>.

⁴³⁹ For a discussion on false positive and false negative variation by demographic groups see National Academics, *Facial Recognition Technology*, cited, 56-59.

⁴⁴⁰ FRT is «often trained mainly on facial images of white men, and much less of women and people belonging to other ethnic groups». FRA, *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, cited, 10. Consequently, FP and FN error rates across demographic groups could be likely reduced by using more diverse and representative data.

⁴⁴¹ *Ibidem*, 4.

⁴⁴² J. CAVAZOS ET AL., *Accuracy Comparison Across Face Recognition Algorithms: Where Are We on Measuring Race Bias?*, in *IEEE Transactions on Biometrics, Behaviour and Identity Science*, 3(1), 2020, 101–111. K.S. KRISHNAPRIYA, K. VANGARA, M.C. KING, V. ALBIERO, K. BOWYER, *Characterizing the Variability in Face Recognition Accuracy Relative to Race*, 2019. Available at: <https://arxiv.org/abs/1904.07325>.

⁴⁴³ P. GROTH, M. NGAN, K. HANAOKA (NIST), *Face Recognition Vendor Test, Part. 3: Demographic Effects*, Report, NISTIR 8280, December 2019. M. NGAN, P. GROTH, *Face Recognition Vendor Test (FRVT): Performance of Automated Gender Classification Algorithms*, NIST, April 2015.

⁴⁴⁴ However, the results might be different in relation to algorithms trained with different datasets (e.g. Asian faces). National Academics, *Facial Recognition Technology*, cited, 57.

⁴⁴⁵ *Ibidem*, 2.

and datasets⁴⁴⁶. These findings brought to light demographic biases embedded in many FRT systems and marked a turning point in the academic and public discourse surrounding the technology⁴⁴⁷. Since then, the issue of unequal algorithmic performance across demographic groups has become a central focus in research on FRT⁴⁴⁸. In relation to the FRT system used by the Metropolitan Police Service (MPS) and South Wales Police, the National Physical Laboratory conducted independent evaluations that identified the poorest performance specifically in the recognition of black female faces. Although the evaluation concluded that the differences in accuracy across demographic groups were not statistically significant overall, the results nonetheless highlighted persistent disparities⁴⁴⁹. In order to adequately take into consideration such discrepancies an Equality Impact Assessment is conducted by the MPS⁴⁵⁰. Higher error rates in FRT for specific demographic groups result in biased outcomes that can ultimately lead to discriminatory practices⁴⁵¹. The implications of biased results and inaccuracies of FRT could be severe: FRT significantly increases the risk of misidentification and members of underrepresented communities are more likely to be wrongfully targeted or investigated⁴⁵², raising profound concerns about racial bias, systemic unfairness, and ethical implications.

⁴⁴⁶ *Ibidem*.

⁴⁴⁷ It must be noted that variations among demographics are ever less evident, since the accuracy of FRT improve rapidly. See *National Academies, Facial Recognition Technology*, cited, 4-5.

⁴⁴⁸ One of the first studies on the topic is J. BUOLAMWINI, T. GEBRU, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, in *Proceedings of Machine Learning Research*, 81, 2018, 1–15. G. PANGELINAN, K.S. KRISHNAPRIYA, V. ALBIERO, G. BEZOLD, K. ZHANG, K. VANGARA, M.C. KING, K.W. BOWYER, *Exploring Causes of Demographic Variations In Face Recognition Accuracy*, 2023. Available at: [arXiv:2304.07175v1](https://arxiv.org/abs/2304.07175v1). M. BRUVERIS, J. GIETEMA, P. MORTAZAVIAN, M. MAHADEVAN, *Reducing Geographic Performance Differential for Face Recognition*, 2020. Available at: [arXiv:2002.12093v1](https://arxiv.org/abs/2002.12093v1). L. HARDESTY, *Study finds gender and skin-type bias in commercial artificial-intelligence systems*, published in *MIT News*, 2018.

⁴⁴⁹ T. MANSFIELD, *Facial Recognition Technology in Law Enforcement Equitability Study Test Strategy*, cited.

⁴⁵⁰ Metropolitan Police Service, *Equality Impact Assessment*. Available at: <https://www.met.police.uk/SysSiteAssets/media/downloads/force-content/met/advice/lfr/impact-assessments/lfr-eia2.pdf>.

⁴⁵¹ CFREU, Article 21; TFEU, Article 10; EU anti-discrimination framework Directive 2000/43/EC, Directive 2000/78/EC, Directive 2006/54/EC, Directive 2004/113/EC, Directive Proposal (COM(2008)462) plays an important role in the European context to limit discrimination based on protected grounds. In algorithmic decision-making processes, however, it seems not to afford sufficient protection for individuals. In T. MADIEGA, H. MILDEBRAT, *Regulating facial recognition in the EU*, cited, 19-21, the authors outline what are the gaps in the EU anti-discrimination framework for adequately address algorithmic discrimination. See also R. XENIDIS, L. SENDEN, *EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination* in ULF BERNITZ ET AL (Eds.), *General Principles of EU law and the EU Digital Order*, Kluwer Law International, 2020.

⁴⁵² See Amnesty international, *Ban dangerous facial recognition technology that amplifies racist policing*, 2021. Similarly, T. MADIEGA, H. MILDEBRAT, *Regulating facial recognition in the EU*, cited, 7; J. LYNCH, *Face Off: Law Enforcement Use of Face Recognition Technology*, published in *Electronic Frontier Foundation*, 12 February 2018. Available at: <https://www.eff.org/wp/law-enforcement-use-face-recognition>. Many articles have been written on FRT's erroneous matches. See K. HILL, *Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match*, published in *The New York Times*, 29 December 2020. Available at: <https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html>. K. JOHNSON, *How Wrongful Arrests Based on AI Derailed 3 Men's Lives*, published in *Wired*, 7 March 2022. Available at: <https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/>. K. HILL, *Facial Recognition Led to Wrongful Arrests. So Detroit Is Making Changes*, published in *The New York Times*, 29 June 2024. Available at: <https://www.nytimes.com/2024/06/29/technology/detroit-facial-recognition-false-arrests.html>.

Biased decisions may also result from the use of other types of AVS. AVS systems used to detect criminal or antisocial behaviours (e.g. assaults, robberies, drug dealing) through specific signals, such as behavioural characteristics (or physical characteristics, such as the presence of a tattoo), can perpetrate the identification and stigmatization of behaviours already perceived as suspicious by the police⁴⁵³. An AVS used to detect “loitering”, for instance, is particularly stigmatizing for homeless people, beggars or sex workers, thus reinforcing the repression of activities that are already subject to police attention. Behavioural data, such as “frequent stops”, “particular crowd movements”, “insufficient or excessive speed”, can trigger the alert and prompt the police to intervene, even if these are ordinary behaviours in public spaces⁴⁵⁴.

To minimize accuracy variations, some authors have distinguished three categories of evaluations that should be carried out in relation to FRT⁴⁵⁵: i) technological evaluations, ii) scenario evaluations and iii) operational evaluations. The purpose of the former is to specifically evaluate the technology behind FRT (i.e. face recognition algorithms). This evaluation defines the performance under specific circumstances, in relation to a standard dataset constructed under controlled circumstances. Since the 1990s, the NIST has conducted one of the most interesting evaluations of face recognition algorithms. In 1993, the Institute started the Face Recognition Technology (FERET) program, whose aim was to sponsor research in face recognition algorithms, collect databases for supporting research and perform their evaluations. In 1999, sponsored by government agencies, the NIST introduced the Face Recognition Vendor Test (FRVT) to evaluate the performance of algorithms used in FRT against specific standards. The FRVT became a global benchmark for FRT performance, and many commercial parties (i.e. vendors) submitted their products for NIST evaluation. In 2017, the NIST introduced a slightly different program, the Face Recognition Technology Evaluation (FRTE) which evaluates the performance of FRT for identification and verification purposes across various criteria. The FRTE extends the scope of application of FRVT to non-commercial parties and includes more tests, such as the one related to demographic effects. As reports published by NIST show, the performance of face recognition algorithms has improved considerably in recent years, especially in ideal conditions. Although this type of evaluation is crucial both for developers, who can improve the technology, and for customers, who can choose between the best performing alternatives, it is not sufficient, since there are many variables in real-life situations that technological evaluations do not consider.

⁴⁵³ AVS may thus encode into technical rules stereotypes and prejudices over specific activities that are already stigmatised, discriminated and repressed by the police.

⁴⁵⁴ See La Quadrature du Net, *Algorithmic Video Surveillance*, cited, 23.

⁴⁵⁵ INTRONA, NISSEBAUM, *Facial Recognition Technology*, cited, 22.

The “scenario evaluation” is directed towards the evaluation of the overall capabilities of the FRT system for a specific application scenario that model a real-world application. Lastly, the purpose of “operational evaluation” is to evaluate an FRT system under real operating conditions. In relation to scenario and operational evaluations, very few information is publicly available. However, evaluations – not only technical ones – are crucial to drive performance improvements, reduce demographic differentials and ensure that FRT is developed and used in accordance with the legal framework and public values. Indeed, the performance of FRT may vary significantly in operational conditions. Furthermore, the lack of evaluations in scenario or operational contexts means that public authorities do not have the necessary information to make informed decisions on which system to purchase and how to use and maintain it appropriately. Experiments under real-life conditions are therefore essential. However, ensuring compliance with the legal framework when carrying out experiments in such circumstances is far from simple⁴⁵⁶.

3.3. The ever-closer relationship between public authorities and private entities

Another concern is the increasingly close relationship between public authorities and private companies in the exercise of surveillance power, as police authorities generally turn to private companies to obtain the technologies and the data with which to supply them⁴⁵⁷. The public-private relationship is multilateral. First, public authorities rely on data collected by private entities⁴⁵⁸. Facial images held by social networks, for example, might be extremely valuable for law enforcement authorities⁴⁵⁹. One of the recent cases that raised global concerns involved a U.S. start-up named Clearview AI⁴⁶⁰, which constructed a database of dozens of billions of facial images from all over the world, extracted from public web sources (social media, online videos) via web scraping, violating platforms’ terms of service, and sold them to many police departments for preventive and repressive purposes. Once this information became public, a class action was led against Clearview AI in the

⁴⁵⁶ As the examples described above show.

⁴⁵⁷ To name a few examples, BigTech companies have some of the most advanced FRT-based systems. It is the case of *Rekognition* developed by Amazon, *Deepface* developed by Facebook and *FaceNet* of Google. In the same terms, MOBILIO, *Tecnologie di riconoscimento facciale*, cited, 20.

⁴⁵⁸ An interesting contribution on this topic is JASSERAND-BREEMAN, *Reprocessing of biometric data for law enforcement purposes*, cited.

⁴⁵⁹ *Ibidem*, 11.

⁴⁶⁰ K. HILL, *The Secretive Company That Might End Privacy as We Know It*, published in *The New York Times*, 18 January 2020. Available at: <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html>. For an interesting analysis of the case from a European perspective see I. NERONI REZENDE, *Facial recognition in police hands: Assessing the ‘Clearview case’ from a European perspective*, in *New Journal of European Criminal Law*, 11(3), 2020, 375.

Northern District Court of Illinois⁴⁶¹ and, similarly, an appeal was filed against the start-up in front of the U.S. District Court for the Eastern District of Virginia. Social networks and internet websites also started warning the company to stop using their content. This case also had considerable repercussions in the European context, since Clearview AI unlawfully collected data from European citizens. In 2022, the Italian DPA, for example, fined the US-based company for EUR 20 million after finding that biometric monitoring techniques have also been applied to individuals in the Italian territory⁴⁶². In the same year, also the Information Commissioner's Office ordered Clearview AI to delete data belonging to UK residents.

Second, technical solutions are generally purchased by public authorities on the market. This is what happened in 2024 Paris Games, where the technical solutions were not developed in-house, but were purchased on the market after a public tender and the contract was awarded to four different private companies⁴⁶³. Purchasing the technical solutions on the market might be problematic because algorithms used for AVS and FRT might be proprietary and hidden by intellectual property rights⁴⁶⁴. It is rather difficult for public authorities to gain access to the details of AVS or FRT's algorithms, data, selection and tuning of parameters. Reliance on technology developed by private companies therefore opens up transparency problems, which prevent public authorities from fully understanding the architecture of FRT, of data used for training, accuracy and error rates of the system.

The issue of ownership of surveillance tools is also crucial to gaining public trust and requires agreements to be made between LEAs and private companies providing the technological infrastructure that precisely establishes obligations of the parties⁴⁶⁵. In addition to transparency concerns, the fact that the design and development of FRTs are predominantly controlled by private companies – primarily motivated by economic interests – raises serious risks to the protection of public values and the safeguarding of fundamental rights. In addition, public data might be accessed by

⁴⁶¹ C. RAMOTTI, *Clearview in USA e privacy: un cambio di paradigma?* published in *Osservatorio sullo Stato Digitale*, IRPA, 15 May 2020.

⁴⁶² Italian Data Protection Authority, *Ordinanza ingiunzione nei confronti di Clearview AI*, 10 February 2022. Available at: <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9751362>. After the news came out, concerns arose in Italy about the possible use of a similar program by public authorities. See V. BONTEMPI, *Un'interrogazione parlamentare sull'uso del riconoscimento facciale in Italia: il caso S.A.R.I.*, published in *Osservatorio sullo Stato Digitale*, IRPA, May 7, 2020.

⁴⁶³ The four companies are: Wintics, Videtics, Chapsvision and Orange Business Service, in partnership with Ipsotek, a subsidiary of Atos (now Eviden).

⁴⁶⁴ Although open-source solutions can be an alternative, they are generally not supported and documented to the level of commercialized ones. National Academies, *Facial Recognition Technology*, cited, 45.

⁴⁶⁵ Y. EZZEDDINE, *Artificial Intelligence in Law Enforcement Surveillance: Citizen Perspectives, Resistance and Counterstrategies*, PhD Thesis, Sheffield Hallam University, 2024. Available at: <https://shura.shu.ac.uk/35469/>. Y. EZZEDDINE, P.S. BAYERL, *Should everyone have access to AI? Perspectives on ownership of AI tools for security*, International Conference on AI Research ICAIR24 Conference Proceedings, 4(1), 2024.

private companies, when they need data to train the model and customize it for the purchasing public authority.

As anticipated, DCNN are the current frontier of FRT. Training DCNNs, as well as deep learning models in general, requires an enormous amount of data. This ever-increasing need for data to train and test face recognition models has led to the creation of increasingly large datasets, along the lines of the Labeled Faces in the Wild⁴⁶⁶. The technique used to create such datasets is generally referred to as “web-scraping”⁴⁶⁷ and allows face images to be collected from online platforms or third-party websites without seeking the consent of individuals to whom those images relate. This practice inevitably poses serious concerns, especially in relation to privacy and data protection⁴⁶⁸.

Considering the high and unacceptable interference of such practices with individuals’ rights and freedoms, the European Union decided to prohibit the “*placing on the market, putting into service for this specific purpose, or the use of AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the Internet or CCTV footage*”⁴⁶⁹. For an AI system to be covered by the ban, four conditions must be met simultaneously: 1) the placing on the market, the putting into service for this specific purpose or the use of an AI system; 2) the purpose must be that of creating or expanding face recognition databases; 3) the databases are created or expanded through AI systems; 4) AI systems are used for the untargeted scraping of images from the internet or CCTV footage⁴⁷⁰. The *Guidelines on prohibited AI practices* specify that only AI systems for the “untargeted scraping” fall into the prohibition. Specifically, the concept of “scraping” refers to

⁴⁶⁶ Examples of datasets are the Oxford’s VGG-Face dataset, Microsoft’s 1M MS Celeb, MegaFace, CASIA WebFace dataset. The literature on the topic is huge. Among others, see Y. GUO, L. ZHANG, Y. HU, X. HE, J. GAO, *Ms-celeb-1m: A dataset and benchmark for large-scale face recognition*, in *European Conference on Computer Vision*, 2016, 87–102. Available at: <https://arxiv.org/abs/1607.08221>. O.M. PARKHI, A. VEDALDI, A. ZISSERMAN ET AL., *Deep face recognition*, Proceedings of the British Machine Vision Conference (BMVC), 2015, 1–12. Available at: <https://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/parkhi15.pdf>. I. KEMELMACHER-SHLIZERMAN, S.M. SEITZ, D. MILLER, E. BROSSARD, *The megaface benchmark: 1 million faces for recognition at scale*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, 4873–4882. D. YI, Z. LEI, S. LIAO, S.Z. LI, *Learning face representation from scratch*, 2014. Available at: <https://arxiv.org/abs/1411.7923>. M. MERLER, N. RATHA, R.S. FERIS, J.R. SMITH, *Diversity in Faces*, 2019, Available at: [arXiv:1901.10436v6](https://arxiv.org/abs/1901.10436v6).

⁴⁶⁷ «Web scraping involves the use of automated software to ‘crawl’ web pages, gather, copy and/or extract information from those pages, and store that information (e.g. in a database) for further use. The information can be anything on a website – images, videos, text, contact details, etc.». Definition given by the ICO. Available at: [Generative AI first call for evidence: The lawful basis for web scraping to train generative AI models | ICO](https://ico.org.uk/for-organisations/articles-and-guidance/privacy-and-data-protection/2018-05-01-generative-ai-first-call-for-evidence-the-lawful-basis-for-web-scraping-to-train-generative-ai-models).

⁴⁶⁸ L. TIEDRICH, *The AI data scraping challenge: How can we proceed responsibly?*, 5 March 2024. Available at: [The AI data scraping challenge: How can we proceed responsibly? - OECD.AI](https://www.oecd.ai/en/web-scraping-do-not-only-poses-risks-for-individuals-privacy-and-data-protection-rights). Web scraping do not only poses risks for individuals’ privacy and data protection rights. Controversies have started in relation to Large Language Model operators who are accessing to scraped data that may include copyrighted materials. On this topic, see Garante per la protezione dei dati personali, *Web scraping ed intelligenza artificiale generativa: nota informativa e possibili azioni di contrasto*, May 2024.

⁴⁶⁹ AI Act, Article 5(1)(e).

⁴⁷⁰ European Commission, *ANNEX to the Communication to the Commission Approval of the content of the draft Communication from the Commission - Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)*, Brussels, 4.2.2025 C(2025) 884 final ANNEX.

the use of different means to “sift through databases and extract information and to make use of that information for another purpose”⁴⁷¹. The concept of “untargeted”, on the other hand, requires scraping to be aimed at the indiscriminate collection of data, without providing for limitations to the data of specific individuals or groups of individuals⁴⁷². Of course, the prohibition does apply when AI systems are involved in the scraping.

Transparency concerns in relation to FRT are quite similar to those highlighted in relation to predictive policing. Regarding tactical obstacles, for instance, public disclosure of FRT systems implemented by police departments may undermine the very rationale of their introduction and hinder law enforcement efforts⁴⁷³. This is one of the reasons why not much public information is currently available on the use of FRT in the European context. As far as technical barriers are concerned, the algorithms used in biometric technologies are generally difficult to understand for a human operator, especially a non-expert one, and in some cases are real black boxes⁴⁷⁴. In addition, the technological barriers discussed above further hinder transparency, consequently hindering public trust.

A crucial aspect to consider is the impact of FRT on fundamental rights⁴⁷⁵. Beyond the right to the protection of personal data, the deployment of FRT – particularly in law enforcement contexts – can affect a broad range of rights. Among others, FRT impacts on the right to privacy – due to the loss of anonymity resulting from constant monitoring and potential identification of individuals⁴⁷⁶ – and the

⁴⁷¹ European Commission, *Commission Guidelines on prohibited artificial intelligence practices*, cited, 78.

⁴⁷² Example provided by the European Commission of a prohibited practice under Article 5(1)(e): «A facial recognition software company collects pictures of faces. The photographs held by the company have been scraped from social media (e.g. Facebook, YouTube, Twitter, Venmo) with an ‘automated image scraper’ that searches the internet and detects images containing human faces. It collects those images with any associated information (such as the source of the image (URL), the geo-localisation, and sometimes the names of the individuals). The facial features are then extracted from the images and transformed into mathematical representations, which are hashed for indexation and future comparison. When a user uploads the image of an individual to the AI system, that system will determine whether that image matches a face in the database. The uploaded image will go through the same mathematical transformation as the scraped images». European Commission, *Commission Guidelines on prohibited artificial intelligence practices*, cited, 79.

⁴⁷³ MADIEGA, MILDEBRAT, *Regulating facial recognition in the EU*, cited, 13.

⁴⁷⁴ *Ibidem*, 20.

⁴⁷⁵ One of the first works on the subject is FRA, *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, cited. On this specific topic see MOBILIO, *Tecnologie di riconoscimento facciale*, cited. In the United States, the relationship between facial recognition and constitutional rights and values has been the subject of study since the early 2000s. *Facial Recognition Technology and Law Enforcement*, Congressional Research Service, September 24, 2020. R.H. THORNBURG, *Face Recognition Technology: The Potential Orwellian Implications and Constitutionality of Current Uses Under the Fourth Amendment*, in *John Marshall Journal of Information Technology & Privacy Law*, 20(2), 2002, 321-346.

⁴⁷⁶ FINOCCHIARO, *Diritto all’anonimato*, cited. Originally, the right to privacy had a negative meaning as a subject’s claim to exclude others from events affecting their personal sphere and encapsulated in the emblematic «right to be let alone». On this specific topic, see L. BRANDEIS, S. WARREN, *The Right to Privacy*, 1890. Only in a second phase did the concept also acquire a different meaning, positive and linked to the principle of self-determination, to be understood as the power to control the circulation of one’s own information. A. WESTIN, *Privacy and Freedom*, in *Administrative Law Review*, 22(1), 1969, 101-106. In the Italian context, the protection of privacy finds a constitutional guarantee in the protection of inviolable rights under Article 2 of the Constitution. S. RODOTÀ, *Persona, riservatezza, identità. Prime note sistematiche sulla protezione dei dati personali*, in *Rivista Critica del Diritto Privato*, Anno XV, 1997. *Idem*, *Privacy, Freedom, and*

right to personal identity – as it reduces individuals to data by transforming their physical characteristics, personality traits and emotional status into quantifiable information⁴⁷⁷.

In addition, due to its pervasive and intrusive nature, the use of FRT in public spaces can have chilling effects on civil and political rights (*i.e.* democratic rights) discouraging individuals from exercising those rights and freedoms guaranteed by both national constitutions and European Charters on fundamental rights (CFREU and ECHR). The perceived sense of constant surveillance⁴⁷⁸ might negatively impact personal freedom⁴⁷⁹, freedom of expression⁴⁸⁰ and freedom of assembly. In ruling *Glukhin v. Russia*⁴⁸¹, the ECtHR specifically found Russia in violation of Articles 8 and 10 of ECHR because of the use of FRT to identify and arrest a political protestor who failed to notify authorities of his intention to hold a solo demonstration. The ECtHR deemed the use of FRT unjustified and intrusive, infringing upon freedom of expression and assembly under Article 10 ECHR. The Court stressed that deploying FRT in public protests can have chilling effects, discouraging individuals from exercising their democratic rights. In addition, the collection and processing of biometric data through FRT was done without sufficient legal safeguards, interfering with the applicant's right to private life (Article 8). The Court concluded that Russian authorities failed to demonstrate that the use of FRT was in accordance with the law, necessary and proportionate. Moreover, there were no clear rules or

Dignity, Conclusive Remarks at the 26th International Conference on Privacy and Personal Data Protection. Available at: <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/1049293#eng>.

⁴⁷⁷ An interesting discussion of the impact of FRT on the right to personal identity can be found in MOBILIO, *Tecnologie di riconoscimento facciale*, cited, 60. In the Italian context the right to personal identity was defined by the Constitutional Court as the « the right to be oneself, understood as respect for the image of participating in associated life, with the acquisitions of ideas and experiences, with the ideological, religious, moral and social convictions that differentiate, and at the same time qualify, the individual». Corte Costituzionale, Judgement No. 13/1994, Par. 5.1, own translation. Alongside personal identity, there is now the concept of digital identity. G. RESTA, *Identità personale e identità digitale*, in *Il Diritto dell'informazione e dell'informatica*, 2007, 511-531. T. PASQUINO, *Identità digitale della persona, diritto all'immagine e reputazione* in E. TOSI (Ed.) *Privacy digitale. Riservatezza e protezione dei dati personali tra GDPR e nuovo Codice Privacy*, Giuffrè, Milano, 2019, 93-114. The right to personal identity can be adversely affected by the use of FRT in several ways. When FRT is employed for verification or authorisation purposes, a false positive – where access is granted to someone other than the authorised individual – constitutes a direct violation of this right. Moreover, the categorization of individuals by FRT systems can lead to partial, reductive, or misleading representations, which may distort how individuals are perceived, undermining the integrity of personal identity.

⁴⁷⁸ An interesting concept is that of «dataveillance» (from «data surveillance»). For a definition of the concept see R.A. CLARKE, *Information Technology and Dataveillance*, in *Communications of ACM*, 31(5), 1988, 498-512. See also J. VAN DIJCK, *Datafication, Dataism and Dataveillance: Big Data Between Scientific Paradigm and Ideology*, in *Surveillance & Society*, 12(2), 2014, 197-208. The Human Rights Council of the United Nations specifically addressed the topic of surveillance through advanced technologies and human rights. See Human Rights Council, *Surveillance and human rights: report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, 2019. Available at: <https://digitallibrary.un.org/record/3814512?v=pdf>. Similarly, National Academies, *Facial Recognition Technology*, cited, 4.

⁴⁷⁹ For a discussion of the possibility of linking FRT and personal freedom, and in what terms, see MOBILIO, *Tecnologie di riconoscimento facciale*, cited, 80.

⁴⁸⁰ K.E. ROY, *Defrosting the Chill: How Facial Recognition Technology Threatens Free Speech*, in *Roger Williams University Law Review*, 27(1), 2022, 185-210.

⁴⁸¹ ECtHR, Judgment of 4 July 2023, Case *Glukhin V. Russia*, Application No. 11519/20. Available at: <https://hudoc.echr.coe.int/eng#%7B%22itemid%22%3A%22001-225655%22%7D>.

transparency regarding the use, storage, or oversight of biometric data processing. Finally, the Court echoed broader concerns that mass deployment of FRT can normalize surveillance, risking systemic rights violations also in democratic societies⁴⁸².

The case shows that in such high-risk scenarios, existing data protection legislation may not be sufficient on its own to address the full range of legal, ethical, and societal concerns. Additional regulatory safeguards and sector-specific rules are necessary to ensure that the deployment of FRT complies with fundamental rights and does not result in disproportionate harm or misuse⁴⁸³. The central issue is not whether FRT should be banned altogether, but rather how to ensure its responsible and proportionate use within a clear and robust legal framework⁴⁸⁴. The goal should be to strike a balance between the potential benefits of FRT – such as enhancing public security – and the need to safeguard fundamental rights. The establishment of a clear and well-defined legal framework governing the use of FRT is essential, as it will significantly influence public perception of these advanced technologies⁴⁸⁵.

3. Function creep in algorithmic policing

Bert-Jaap Koops, in the foundational conceptual treatment of the subject, defines function creep as “*something’s function [which] is moving beyond its original purpose in a way that was apparently unforeseen by its developers, users, or the public*”, in a slow, gradual and imperceptible way⁴⁸⁶. Three elements distinguish function creep from benign technological evolution. First, *imperceptibility*: the change occurs gradually, making it difficult to identify a precise moment of transformation. Second, *transformativeness*: the change fundamentally alters the system’s function rather than merely extending it. Third, *contestability*: the change raises legitimate normative objections concerning its legality, proportionality, or democratic authorisation.

⁴⁸² K.A. GATES, *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance*, NYU Press, 2011. D. LYON, *La società sorvegliata. Tecnologie di controllo della vita quotidiana*, Feltrinelli, 2003.

⁴⁸³ Consultative Committee of Convention 108, *Guidelines on facial recognition*, cited.

⁴⁸⁴ In GIKAY, *Regulating use by law enforcement authorities of live facial recognition technology in public spaces*, cited, the author proposes an incremental approach. Considering the impacts that the use of FRT could have on democratic rights, the author proposes to introduce legal restrictions on the use of FRT in contexts of critical civil and political activities. By contrast, the use of FRT in non-political contexts (such as in large scale events) where a security threat does exist should be allowed, although under defined conditions. «*These kind of restrictions should not take the form of a blanket prohibition, as a proportionality assessment needs to be conducted in the given circumstances*». See National Academies, *Facial Recognition Technologies*, cited, 6.

⁴⁸⁵ Trust is widely recognised as a critical factor in securing public acceptance, and transparent, accountable regulation plays a key role in building that trust. It clearly emerges from the work EZZEDDINE, *Artificial Intelligence in Law Enforcement Surveillance*, cited.

⁴⁸⁶ B.-J. KOOPS, *The Concept of Function Creep*, in *Law, Innovation and Technology*, 13(1), 2021, 37.

In the algorithmic policing context, function creep assumes a distinctive character because the technical architecture of data-driven systems actively facilitates this expansion. Unlike analogue surveillance instruments with inherent physical or operational limits, algorithmic systems are designed for interoperability, scalability, and pattern extraction across heterogeneous datasets. Arif Kornweitz has argued, provocatively, that function creep is better understood as the *modus operandi* of algorithmic systems, a testament to their technical mutability and political entanglement, challenging the standard framing that treats function creep as an exceptional deviation from intended use⁴⁸⁷ Kornweitz suggests that algorithmic policing systems do not merely risk function creep; they are structurally predisposed to it, because their operational value increases precisely as the scope of data inputs and analytical outputs expands.

Function creep in algorithmic policing manifests through distinct but interconnected vectors. Purpose expansion, for example, occurs where systems deployed for serious crime investigation are gradually applied to minor offences. Population expansion describes the extension of targeted tools to encompass entire communities. Data source proliferation involves the integration of datasets from multiple and disparate sources. Temporal extension arises where systems designed for post-event analysis are repurposed for real-time or predictive use. Finally, institutional migration refers to technology transferring from the private to the public sector, or across government agencies, without adequate legal bases. Each of these vectors is traceable in the case studies examined above.

The predictive policing case study reveals function creep operating across all five vectors identified above. For instance, the Dutch SyRI system combined data from employment, tax, social security, immigration, and housing databases to generate risk scores for welfare and tax fraud. SyRI exemplified data source proliferation and population expansion: data collected across multiple government agencies under separate legal bases was aggregated into a profiling tool applied to entire communities rather than individual suspects. The District Court of The Hague struck down SyRI in February 2020⁴⁸⁸, holding that the SyRI legislation – Article 65 of the SUWI Act and Chapter 5a of the SUWI Decree – did not strike a fair balance between fraud prevention and individuals’ right to privacy under Article 8 ECHR.. Germany’s hessenDATA, built on Palantir’s Gotham platform, demonstrated purpose expansion and temporal extension. Originally deployed in Hesse in 2017 for counterterrorism and organised crime investigations, its operational scope expanded progressively to encompass minor offences, with the system ultimately used thousands of times annually for routine police work. The

⁴⁸⁷ A. KORNWEITZ, *A New AI Lexicon: Function Creep*, AI Now Institute, 4 August 2021. Available at: <https://ainowinstitute.org/publications/collection/a-new-ai-lexicon-function-creep>.

⁴⁸⁸ District Court of The Hague, 5 February 2020, ECLI:NL:RBDHA:2020:865.

German Federal Constitutional Court ruled on 16 February 2023⁴⁸⁹ that the Hessian and Hamburg provisions authorising automated data analysis were unconstitutional, violating the right to informational self-determination under Article 2(1) read with Article 1(1) of the Basic Law. The Court held that automated data analysis constitutes a separate interference with fundamental rights, particularly serious where processing generates machine-driven evaluations of a predictive policing character. The decision required that legislation authorising the processing contain a set of procedural safeguards directly targeting the incremental, undocumented expansion characteristic of function creep. The Chicago Strategic Subject List illustrates population expansion in its most extreme form. Initially targeting approximately 400 high-risk individuals, the system eventually encompassed almost 400.000 persons. This ten-thousand-fold expansion occurred without legislative authorisation or public deliberation.

The face recognition case study also reveals function creep dynamics. The technology's capacity to identify individuals at a distance, without their knowledge or cooperation, creates what the EDPB and EDPS described in Joint Opinion 5/2021 as potentially marking "*the end of anonymity in publicly accessible spaces*"⁴⁹⁰. The Clearview AI model exemplifies private-to-public sector creep in its purest form. As of October 2021, the private company scraped over ten billion facial images from the internet to build a recognition database sold primarily to law enforcement authorities. CEO Hoan Ton-That told the BBC in March 2023 the database contained 30 billion images, and later company statements referenced over 50 billion⁴⁹¹. The Italian SARI Real Time system demonstrates expansion from targeted to mass surveillance. In its April 2021 opinion, the Italian DPA identified the core function creep dynamic: while the system was presented as targeting specific individuals, the identification of a person would be achieved through the processing of biometric data of all those present in the monitored space, shifting from targeted surveillance of particular individuals to the possibility of universal surveillance⁴⁹². The watchlist could store up to 10,000 faces, and there were no regulatory criteria governing who could be placed on it. *R (Bridges) v Chief Constable of South Wales Police*, the world's first successful legal challenge to police face recognition, identified function creep risks in watchlist management⁴⁹³. The Court of Appeal found that South Wales Police's use of AFR Locate was unlawful because the legal framework had "*fundamental deficiencies*": there was "*no clear guidance on who could be put on a watchlist*" and the system afforded "*too broad a discretion*" to

⁴⁸⁹ German Federal Constitutional Court, 16 February 2023, 1 BvR 1547/19; 1 BvR 2634/20.

⁴⁹⁰ EDPB and EDPS, Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (AI Act).

⁴⁹¹ J. CLAYTON, B. DERICO, *Clearview AI used nearly 1m times by US police, it tells the BBC*, 28 March 2023. Available at: <https://www.bbc.com/news/technology-65057011>.

⁴⁹² Garante per la protezione dei dati personali, Opinion on the SARI Real Time facial recognition system, April 2021.

⁴⁹³ *R (Bridges) v Chief Constable of South Wales Police* [2020] EWCA Civ 1058.

individual police officers. Over fifty deployments between 2017 and 2019 scanned approximately 500,000 faces, the overwhelming majority belonging to individuals not suspected of any wrongdoing. The decision resonates with the ECtHR's holding in *Catt v United Kingdom*,⁴⁹⁴ where police retention of sixty-six database entries recording a 94-year-old peace activist's attendance at protests was held to violate Article 8 ECHR. *Catt* represents paradigmatic function creep: intelligence infrastructure built for monitoring violent protest groups extended to encompass peaceful political activity.

European law deploys two primary legal mechanisms against function creep in algorithmic policing: the purpose limitation principle, operationalised through the LED, and the fundamental rights framework of necessity and proportionality under the ECHR and the CFREU. As far as the former is concerned, the LED requires that personal data be “*collected for specified, explicit and legitimate purposes and not processed in a manner that is incompatible with those purposes*”⁴⁹⁵. However, the LED's implementation of this principle is significantly weaker than the GDPR's equivalent. In fact, Article 4(2) LED permits processing for “*other purposes*” within the broad scope of Article 1(1) – encompassing prevention, investigation, detection, prosecution of criminal offences, and execution of criminal penalties – provided the controller is authorised by law and the processing is necessary and proportionate. The CJEU confirmed that even shifting from detection or investigation to prosecution of the same offence constitutes processing for a “*different purpose*” requiring compliance with Article 4(2)⁴⁹⁶. Nevertheless, the breadth of permissible law enforcement purposes, combined with the simplified compatibility test, creates significant latitude for incremental purpose expansion. Moreover, the LED's Article 6 requirement to distinguish between categories of data subjects – namely, suspects, convicted persons, victims, witnesses, and third parties – directly targets a key function creep mechanism. Predictive policing systems, however, inherently blur these categories: when algorithms score individuals, they do not differentiate between suspects and innocent bystanders. The qualifying phrase “*as far as possible*” in Article 6 further weakens its constraining force for systems incapable of making such distinctions.

The central question thus is whether purpose limitation can be adapted to the algorithmic context, or it requires fundamental reconceptualization. As law enforcement authorities across the European Union and beyond deploy increasingly sophisticated algorithmic systems, from predictive policing tools to biometric surveillance infrastructure, the practical adequacy of purpose limitation as a regulatory instrument is subject to mounting challenge. Such technologies continuously generate,

⁴⁹⁴ ECtHR, Judgement of 24 January 2019, *Catt v. United Kingdom*, Application No. 43514/15.

⁴⁹⁵ LED, Article 4(1)(b).

⁴⁹⁶ CJEU, Judgement of 8 December 2022, *VS v Inspektor v Inspektorata kam Visshia sadeben savet*, Case C-180/21, ECLI:EU:C:2022:967.

collect, and process data for emergent purposes that cannot be specified in advance⁴⁹⁷. As Mireille Hildebrandt states, this challenge is structural rather than incidental, being the defining characteristic of ambient intelligence its capacity to produce new inferences from data combinations that were never anticipated at the point of collection. However, purpose specification, conceived as a one-time obligation imposed at the moment of data acquisition, is constitutively inadequate as a regulatory response to technologies whose operational outputs are, by design, unpredictable. Hildebrandt's proposed solution, namely "*legal protection by design*", calls for embedding legal safeguards into the technological architecture itself. Her argument is that traditional *ex post* remedies are inadequate for systems operating in real-time and at scale: by the time a data subject becomes aware of a violation, the harms may be irreversible and the causal chain obscured. This reconceptualisation moves beyond treating purpose limitation as a data controller obligation and reimagines it as a design constraint, requiring, for instance, that predictive policing algorithms be technically incapable of processing data beyond their validated operational scope. This proposal has significant institutional implications. If purpose limitation is reconceived as a design requirement, regulatory oversight necessarily shifts upstream, to the stage of system procurement and certification. This proposed solution closely mirrors the architecture of the AI Act, as will be examined in Chapter 3.

The necessity and proportionality framework established under Articles 7 and 8 of the CFREU and Article 8 of the ECHR provides a constitutionally grounded constraint on such deployments, providing the standards most directly applicable to algorithmic law enforcement. The CJEU's data retention jurisprudence has established that indiscriminate data collection constitutes a "*particularly serious*" interference with fundamental rights requiring strict justification. In *Digital Rights Ireland*⁴⁹⁸, the Court invalidated the Data Retention Directive on the basis that it failed to limit retention by "*objective criteria*" linking it to specific threats. This decision established the foundational proportionality standard: general and indiscriminate retention of communications data is, in principle, incompatible with Articles 7 and 8 of the CFREU. The Court's reasoning rested on the absence of any differentiation between categories of persons, the lack of a causal nexus to serious crime, and the failure to require independent prior review of access. *Tele2 Sverige/Watson*⁴⁹⁹ extended the Digital Rights Ireland standard to national legislation, requiring that retention be targeted (limited by categories of data, persons, and temporal scope) and that access be subject to prior independent judicial review. The Court

⁴⁹⁷ M. HILDEBRANDT, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*, Edward Elgar, 2015.

⁴⁹⁸ CJEU, Grand Chamber, Judgement of 8 April 2014, *Digital Rights Ireland Ltd v. Minister for Communications*, Joined Cases C-293/12 and C-594/12, ECLI:EU:C:2014:238.

⁴⁹⁹ CJEU, Grand Chamber, Judgement of 21 December 2016, *Tele2 Sverige AB v. Post-och telestyrelsen; Secretary of State for the Home Department v. Watson and Others*, Joined Cases C-203/15 and C-698/15, ECLI:EU:C:2016:970.

clarified that national schemes permitting general and indiscriminate access to retained data are incompatible with EU law. It is noteworthy that *La Quadrature du Net*⁵⁰⁰ introduced a significant qualification to the strict prohibition established in earlier cases, carving out exceptions for genuine and present or foreseeable national security threats. This doctrinal development creates a structural risk of what may be characterised as cross-regime function creep: data retained under the national security exception may be leveraged for ordinary law enforcement purposes. This is precisely the dynamic that the LED's purpose limitation principle is designed to prevent, yet may be structurally unable to contain. *Ligue des Droits Humains*⁵⁰¹ provides the most directly applicable precedent for algorithmic policing. Scrutinising the Passenger Name Record (PNR) Directive, which involves mass automated profiling of airline passengers, the Court articulated three holdings of particular significance. First, algorithmic screening must employ pre-determined criteria that are “*specific, reliable, and non-discriminatory*”. This requirement effectively imposes a *de facto* prohibition on opaque machine learning algorithms deployed for mass screening purposes. Second, all positive matches must be subjected to individual, non-automated review by a competent authority, preserving a human-in-the-loop safeguard that directly constrains fully automated decision-making in law enforcement contexts. Third, full data retention must be limited to six months, rather than the Directive's five-year retention period. These standards, if transposed to predictive policing and face recognition regimes, would substantially constrain the discretionary, opaque, and temporally unlimited data processing that characterises current deployments.

As far as the ECtHR's jurisprudence is concerned, *S. and Marper v. United Kingdom*⁵⁰² established the foundational principle that blanket and indiscriminate retention of biometric data of unconvicted persons violates Article 8 ECHR. The Court articulated a requirement for “*clear, detailed rules governing scope and application of retention measures*” incorporating minimum safeguards regarding duration, storage, usage, third-party access, and procedures for destruction. These requirements directly constrain the indefinite data pooling that enables algorithmic systems to repurpose investigative data for predictive analytics. Notably, the Court observed the over-representation of ethnic minorities in the United Kingdom's DNA databases, thereby anticipating the discriminatory dynamics subsequently documented in empirical studies of predictive policing systems. *Catt v. United Kingdom*⁵⁰³ extended the principles established in *S. and Marper* to police intelligence databases. The

⁵⁰⁰ CJEU, *La Quadrature du Net and Others*, cited.

⁵⁰¹ CJEU, Grand Chamber, Judgement of 21 June 2022, *Ligue des Droits Humains v. Conseil des ministres*, Case C-817/19, ECLI:EU:C:2022:491.

⁵⁰² ECtHR Grand Chamber, Judgement of 4 December 2008, *S. and Marper v. United Kingdom*, Application Nos. 30562/04 and 30566/04.

⁵⁰³ *Catt v. United Kingdom*, cited.

Court found that indefinite retention without meaningful review was disproportionate even where initial collection was lawfully justified. Significantly, it rejected the Government's contention that database review would be administratively burdensome, stating that it would be contrary to Article 8 if the Government could create a database in such a manner that the data in it could not be easily reviewed or edited.

The foregoing analysis reveals a converging body of constitutional doctrine applicable to algorithmic policing. From the CJEU's data retention line and the ECtHR's biometric data jurisprudence, the following doctrinal standards may be distilled: (i) the prohibition on indiscriminate data collection without objective, specific, and non-discriminatory criteria; (ii) the requirement for independent prior review of access to retained data; (iii) the imposition of temporal limits on retention, with structured review mechanisms; (iv) the prohibition on fully automated decision-making without human review of positive matches; and (v) heightened protection for sensitive categories of data.

While the necessity and proportionality framework under the CFREU and the ECHR, together with the purpose limitation regime of the LED, provide significant constitutional and legislative constraints on algorithmic policing, this analysis must be situated within the broader regulatory landscape that has emerged since 2024. The entry into force of Regulation (EU) 2024/1689 introduces a distinct and complementary layer of obligations directly applicable to algorithmic law enforcement systems. The AI Act imposes binding requirements that both reinforce and, in certain respects, extend beyond the protections afforded by existing data protection instruments, and accordingly merits sustained consideration in any assessment of the regulatory framework governing algorithmic policing.

CHAPTER 3

The regulation of AI systems for law enforcement purposes within the AI Act

1. Overview of the Chapter	189
Part I: the Artificial Intelligence Act.....	192
1. The European path towards the regulation of AI.....	192
2. The first binding rules on AI: the European Union's Artificial Intelligence Act.....	194
3. The complex legal architecture of the AI Act.....	205
3.1. The horizontal approach.....	206
3.2. The risk-based approach.....	209
3.3. The by-design regulation approach	218
3.4. The co-regulatory approach: the technical standards	222
3.5. Considerations on the overall AI Act's structure	228
Part II: Law enforcement in the AI Act	232
1. The AI Act's provisions in the law enforcement context	232
2. Predictive Policing in the AI Act.....	232
2.1. Prohibited predictive policing	236
2.2. Other categories of predictive policing	242
3. Automated video surveillance in the AI Act.....	245
3.1. AI systems for categorising natural persons.....	246
3.2. Biometric identification systems	250
3.2.1. Exceptions to the prohibition: safeguards.....	256
3.2.2. High-risk biometric identification systems.....	260
3.3. Biometric verification	261
4. Considerations on the established legal framework.....	263

1. Overview of the Chapter

The two applications discussed in Chapter 2 illustrate how AI equips public authorities with innovative instruments that can significantly transform the exercise of public power¹. By incorporating AI into public functions, public authorities do not merely gain efficiency; they also reshape the dynamics of

¹ On this topic see B. MARCHETTI, *Artificial Intelligence and Public Authorities: Does the European AI Act Protect Public Values?*, in *European Review of Public Law*, 36(1), 2024, 67-90.

power, including the relationship between public institutions and private entities that generally design, supply, or manage these technological solutions. Currently, AI emerges not as a substitute for human decision-making but as a complementary tool²: it supports human operators by providing data-driven insights, structured knowledge, and predictive analyses that can inform and guide decisions. In this way, the technology operates as a decision-support mechanism, while the ultimate responsibility for the decision remains vested in the human authority. However, as analysed in Chapter 2, the integration of AI into public decision-making processes is not free from risks. These risks include, but are not limited to, the possibility of discriminatory outcomes resulting from biased datasets or flawed algorithms, opacity in the functioning of AI systems that compromises transparency, a potential reduction in accountability when responsibility becomes diffused between human and machine, and concerns over privacy stemming from the large-scale use of personal data³. Such risks, if not properly addressed, may undermine the very legitimacy of public action.

Therefore, while AI presents significant opportunities, its deployment by public bodies must be carefully aligned with core public values. Public authorities have a duty to ensure that AI systems operate in a manner consistent with fundamental principles, including transparency of administrative action, equality of treatment, protection of personal data, and the safeguarding of fundamental rights⁴. Since public values are not intrinsically embedded in the design and functioning of AI systems – whose

² *Ibidem*, 71.

³ The literature on this topic is huge. Among others, see J.V. FERNANDEZ, *Artificial Intelligence in Government: Risks and Challenges of Algorithmic Governance in the Administrative State*, in *Indiana Journal of Global Legal Studies*, 30(1), 2023, 65. A. RACHOVITSA, N. JOHANN, *The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons learned from the Dutch SyRI Case*, in *Human Rights Law Review*, 22(2), 2022, 1-15. For the Italian debate on the topic, see G. AVANZINI, *Decisioni amministrative e algoritmi informatici. Predeterminazione analisi predittiva e nuove forme di intellegibilità*, Editoriale scientifica, 2019; A. SIMONCINI, *Profili costituzionali della amministrazione algoritmica*, in *Rivista trimestrale di diritto pubblico*, 4, 2019, 1149-1188; J.-B. AUBY, *Il diritto amministrativo di fronte alle sfide digitali*, in *Istituzioni del Federalismo*, 3, 2019, 619-641; M. BASSINI, L. LIGUORI, O. POLLICINO, *Sistemi di intelligenza artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, in F. PIZZETTI (Ed.) *Intelligenza artificiale, protezione dei dati personali e regolazione*, Giappichelli, 2018; C. CASONATO, *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, in *Biolaw Journal*, Special Issue, 2, 2019, 711-725; C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Diritto pubblico comparato ed europeo*, Special Number, 2019, 101-130; F. COSTANTINO, *Rischi e opportunità del ricorso delle amministrazioni alle predizioni dei Big Data*, in *Diritto pubblico*, 1, 2019, 43-70; S. CIVITARESE MATTEUCCI, *Umano troppo umano. Decisioni amministrative automatizzate e principio di legalità*, in *Diritto pubblico*, 1, 2019, 5-41; A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal*, 1, 2019, 63-89; D.U. GALETTA, J.G. CORVALÁN, *Intelligenza artificiale per una pubblica amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *Federalismi.it*, 3, 2019, 2-23.

⁴ The concept of “fundamental rights” must be distinguished from the concept of “human rights”. Human rights are universal in nature, while fundamental rights, although closely related and largely overlapping with human rights, are framed within specific legal and constitutional contexts. Therefore, their scope and protection depend on *if* and *how* human rights are translated into binding norms within a particular legal order. In the European context, for example, fundamental rights are enshrined in the CFREU and in the constitutions of the Member States. As the European Union Agency for Fundamental Rights explains, the term *fundamental rights* is generally used within the EU to denote the concept of human rights in its internal legal and constitutional setting, whereas the broader term *human rights* is more typical of international law. However, the substance of the two notions is largely equivalent, as illustrated by the strong parallels between the CFREU and the ECHR. See FRA, *What are fundamental rights?*. Available at: <https://fra.europa.eu/en/content/what-are-fundamental-rights>.

CHAPTER 3

development is frequently driven by considerations of efficiency, accuracy, output optimisation, and private economic interests – the European Union has opted for a regulatory approach aimed at mitigating and managing the risks associated with AI systems and ensuring that their placing on the market, putting into service and deployment remains compatible with fundamental rights and with the core values underpinning the European legal order.

Against this background, the first part of the Chapter (Part I) examines the Artificial Intelligence Act from a general perspective, analysing its overall structure, regulatory logic, and guiding principles. The second part of the Chapter (Part II) focuses specifically on the provisions of the AI Act applicable to the law enforcement sector, with the objective of assessing whether the Regulation establishes adequate and effective safeguards for the deployment of AI systems in this particularly sensitive domain.

Part I: the Artificial Intelligence Act

1. The European path towards the regulation of AI

Humans and technologies are deeply intertwined: on one hand, technologies are created by humans and inevitably reflect their social norms, values, and broader economic, political, and cultural contexts; on the other hand, technologies influence how people work, move, communicate, and live. Since the beginning of the past decade, it had become clear that digital technologies, which were becoming an ever more central part of people's lives, were transforming the world at an unprecedented pace, impacting on the individual and the society as a whole⁵.

To harness the opportunities offered by digital technologies, in 2015 the European Union introduced the Digital Single Market strategy⁶, as a *species* of the broader Single Market strategy⁷. From the beginning, the strategy focused on removing barriers within the digital sector, while preserving, at the same time, fundamental rights and freedoms of individuals⁸. This highlighted the need for a robust regulatory framework to ensure that technological development aligns with European values. As a result, digital regulation became a key pillar of the European Commission's strategic agenda⁹.

⁵ See European Commission, *White Paper: On Artificial Intelligence - A European approach to excellence and trust*, Brussels, 19.2.2020, COM(2020) 65 final, 2.

⁶ The Digital Single Market strategy was adopted in May 2015 and it represented one of the European Commission's political priorities. European Commission, *Communication. A Digital Single Market Strategy for Europe*, Brussels, 6.5.2015, COM(2015) 192 final. See K. NYMAN METCALF, I. PAPAGEORGIOU, *The European Union Digital Single Market—Challenges and Impact for the EU Neighbourhood States*, in *Baltic Journal of European Studies*, 8(2), 2018, 7-23. The Digital Single Market Strategy became a priority in the Commission's Agenda 2019-2024 and 2024-2029, under the Presidency of Ursula Von der Leyen. In 2022 the European Commission published the *Digital Compass 2030* which, among other aspects, underlines the importance of putting people and their rights at the center of the digital transformation. European Commission, *Digital Compass 2030, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*, Brussels, 9.3.2021, COM(2021) 118 final, 2.

⁷ TFEU, Article 26, paragraph 2, defines the Single Market as “[...] an area without internal frontiers in which the free movement of goods, persons, services and capital is ensured in accordance with the provisions of the Treaties”. It is well known that over the years there has been a linguistic change from “common market” to “single market” and finally to “internal market”. Despite the change in terminology, the Court of Justice has emphasised that these adjectives are considered interchangeable. K. MORTELMANS, *The Common Market, the Internal Market and the Single Market, What's in a Market?*, in CMLR, 1998. S. WEATHERILL, *The Internal Market as a Legal Concept*, Oxford, 2016.

⁸ The strategy is based on three main pillars: 1) improving access to digital goods and services, 2) an environment where digital networks and services can prosper and 3) digital as a driver for growth.

⁹ Until 2015 the digital regulations were mainly Directive 95/46/EC, Directive 2000/31/EC and Directive 2013/37/EU. After 2015 multiple legislative acts have been adopted: the GDPR in 2016, the ePrivacy Regulation and Portability Regulation in 2017; the Geo-blocking Regulation in 2018, the Digital Content Directive and the Copyright Directive in 2019, the eIDAS Regulation in 2021, the Data Act, the Digital Markets Act, the Digital Operational Resilience Act, the Network Information Security Directive, the Information Security Regulation, the Data Governance Act and the Digital Services Act in 2022, the e-Evidence package in 2023 and the AI Act in 2024. Another interesting legislative act was the AI Liability Directive, but the European legislator eventually decided to withdraw it. European Commission, *Annexes to the communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*, Commission work programme 2025, Strasbourg, 11.2.2025, COM(2025) 45

However, regulating the digital realm has proven to be an exceptionally complex challenge¹⁰: the rapid pace of technological advancement often outstrips the slower processes of law-making; crafting effective rules also requires an interdisciplinary approach that reconciles legal aspects with technological feasibility; moreover, traditional legal systems grounded in national jurisdictions struggle to adequately address violations by global tech giants¹¹.

A few years after the launch of the Digital Single Market strategy, AI has become a central issue, and its regulation has become one of the European Union's priorities¹². It is generally acknowledged that AI is not a simple technology, rather it is a "transformative technology"¹³, because it has proven to entail – due to its opacity, complexity, continuous adaptability and unpredictability, autonomy and functional dependency on data¹⁴ – multiple transformations¹⁵. AI opens new opportunities, maintaining and enhancing human actions, but it can also be used in ways that are incompatible with the realisation of human values, restricting human freedoms and hindering human capabilities¹⁶. As it has been

final, 26. On the specific topic of EU digital regulation see A. VOLPATO, *The publicity of EU law and the privatization of EU digital regulation*, in *Maastricht Journal of European and Comparative Law*, 31(3), 2024, 319-332. Available at: <https://doi.org/10.1177/1023263X241301937>.

¹⁰ In these terms, see A. ADINOLFI, *Evoluzione tecnologica e diritti fondamentali: qualche considerazione sulle attuali strategie normative dell'Unione*, in *I Post di AISDUE*, 2023, 15. A. BRADFORD, *The Expansion of Empires*, in A. BRADFORD, *Digital Empires. The Global Battle to Regulate Technology*, Oxford University Press, 2023, 324-325.

¹¹ Tech giants such as Google, Meta, and Apple are increasingly regarded as *quasi-state actors*, wielding influence that rivals or even surpasses that of many governments. See L. TORCHIA, *Lo Stato digitale. Una introduzione*, il Mulino, 2025.

¹² The first time the topic of AI entered the political discourse was in 2017, when the European Union emphasised the importance of promoting investment and research on this topic. After this moment, a plethora of policy documents followed. In 2018 the European Commission adopted a first Communication. European Commission, *Communication From The Commission To The European Parliament, The European Council, The Council, The European Economic And Social Committee And The Committee Of The Regions Artificial Intelligence for Europe*, Brussels, 25.4.2018 COM(2018) 237 final. In the same year, the High Level Expert Group on Artificial Intelligence was set up by the European Commission and in 2019 the group published the Ethics Guidelines for Trustworthy AI. In 2020 the European Commission published a White Paper on Artificial Intelligence. European Commission, *White Paper. On Artificial Intelligence - A European approach to excellence and trust*, Brussels, 19.2.2020, COM/2020/65 final. In 2021 the European Commission published another Communication and the Proposal for a regulation of AI. European Communication, *Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions Empty Fostering A European Approach To Artificial Intelligence*, Brussels, 21.4.2021, COM(2021) 205 final; European Communication, *Proposal For A Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts*, Brussels, 21.4.2021, COM/2021/206 final. For an interesting discussion on the regulation of AI see A. ADINOLFI, *L'Unione europea dinanzi allo sviluppo dell'intelligenza artificiale: la costruzione di uno schema di regolamentazione europeo tra mercato unico digitale e tutela dei diritti fondamentali*, in S. DORIGO (Ed.), *Il ragionamento giuridico nell'era dell'intelligenza artificiale*, 2020.

¹³ ROSS GRUETZEMACHER, JESS WHITTLESTONE, *The transformative potential of artificial intelligence*, in *Futures*, 135, 2022, 1-11.

¹⁴ These characteristics were identified by the European Commission in the Impact Assessment accompanying the AI Act. European Commission, *Commission Staff Working Document Impact Assessment - Accompanying the Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts*, Brussels, 21.4.202, SWD(2021) 84 final, COM(2021) 206 final, 28.

¹⁵ Among others, see B. RASHID, A. K. KAUSIK, *AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications*, in *Hybrid Advances*, 7, 2024, Article 100277.

¹⁶ Multiple uses of AI can be identified : a) "correct use" can create opportunities in multiple sectors; b) "underuse", due to a failure to make the most appropriate decisions or to fear and scepticism towards AI, can create opportunity costs; c)

previously observed, AI entails several potential risks, such as opaque decision-making, gender-based or race-based discrimination, non-compliant processing of personal data. The above considerations have called for a united action at the European level in order to exploit the opportunities it promises and address the risks it poses, while ensuring the correct functioning of the internal market of AI systems. In the following sections, this chapter begins to address the central research question, namely whether public values are adequately protected by the legal system when law enforcement activities are enhanced through the use of AI systems. To this end, it is first necessary to analyse the architecture of the AI Act in order to assess its strengths and limitations in safeguarding fundamental rights and public values.

2. The first binding rules on AI: the European Union’s Artificial Intelligence Act

First and foremost, when seeking to regulate a particular domain, phenomenon, or activity, it is essential to define the *object* of the regulation and the *type* of the regulatory approach to be adopted. As for the *object* of the regulation, I analysed in Chapter 1 that the European legislator did not aim to regulate AI, rather to set high standards for AI systems that are to be placed on the EU market¹⁷. Regarding the *type* of regulation to be employed, when it comes to technology, more than relying on a single legal instrument, it is essential to establish a “*coordinated system of governance*” where multiple regulatory instruments are taken into consideration¹⁸.

In general, different regulatory approaches can be distinguished based on either the *effectiveness* of the rules or the *entity* issuing them. Regarding the *effectiveness* of the rules, the traditional distinction is between hard law and soft law. Hard law refers to rules that, from a legal standpoint, are binding and, if violated, are subject to judicial enforcement and compulsory execution. In general, it encompasses all rules enacted by national or European legislative bodies. Soft law, on the other hand, refers to non-binding instruments which includes preparatory or informative acts (such as white papers and green

incorrect or excessive use (“misuse” or “overuse”) can bring with it dangerous risks. Carlo Casonato correctly argues that AI can be understood “*as both the greatest threat to human rights and human survival itself, and as the lever for a formidable advancement of mankind in every sphere*”. C. CASONATO, *Per una intelligenza artificiale costituzionalmente orientata* in D’ALOIA (Ed.), *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, FrancoAngeli, 2020, 132.

¹⁷ While defining what is meant by “AI” remains complicated and identifying precise boundaries for the concept is virtually impossible, regulating the uses of AI, taking into account the consequences they have on the rights and safety of European citizens, is not only possible but also essential. In this sense, F. ROSSI, *A European approach to the regulation of artificial intelligence*. Online meeting held on 23 April 2021 organised by CEPS. Available at: <https://www.ceps.eu/ceps-events/a-european-approach-to-the-regulation-of-artificial-intelligence/>.

¹⁸ The idea of a “coordinated system of governance” for the various regulatory options is set out in ANDREA SIMONCINI, *Verso la regolamentazione della Intelligenza Artificiale. Dimensioni e governo*, in *BioLaw Journal*, 2, 2021, 414. In this paper, the author emphasises that such an option is justified because the more traditional forms of regulation have shown serious shortcomings in terms of effectiveness over the years.

papers, action plans, and programs), interpretative acts (such as communications), conclusions, declarations, resolutions, and finally, recommendations and opinions. These are non-prescriptive rules which, while not judicially enforceable, carry a certain degree of persuasiveness and ability to guide and orient their recipients¹⁹.

About the *entity* issuing the rules, a distinction is generally made between straight regulation, self-regulation, and co-regulation²⁰. Straight regulation refers to a set of rules established by public bodies or authorities (such as international organizations, European institutions, national Parliaments, etc.) that are external to the individuals or entities subject to those rules. The defining feature of this type of regulation is that it is imposed from the outside, often through binding legal instruments and backed by enforcement mechanisms. In this case, straight regulation also falls within the category of hard law. By contrast, private self-regulation occurs when economic actors, as an expression of their private autonomy, set the rules they commit to follow in their mutual relations and in dealings with third parties in the marketplace²¹. The term co-regulation refers to a form of stakeholder regulation that is promoted, guided, or supervised by a third party (whether an official body or an independent regulatory authority), typically equipped with review, monitoring, and in some cases sanctioning powers²². Therefore, while the rules are created by private entities or their associations, these same rules are subject to indirect oversight by a public authority that ensures their compliance. Historically, the term “regulation” referred mainly, if not exclusively, to straight-hard law. However, today, private self-regulation and co-regulation are considered important complementary or supplementary tools that fall within this concept²³.

¹⁹ O. STEFAN, M. AVBELJ, M. ELIANTONIO, M. HARTLAPP, E. KORKEA-AHO, N. RUBIO, *EU Soft Law in the EU Legal Order: A Literature Review*, in *King's College London Law School Research Paper*, 9, 2019.

²⁰ For the following discussion reference is made to European Economic and Social Committee, *Opinion of the European Economic and Social Committee on Self-regulation and co-regulation in the Community legislative framework*, 2015/C 291/05, 4 September 2015, paragraph 4.5 letter d). See also L.A.J. SENDEN, *Soft Law, Self Regulation and Co-regulation in European Law: Where do they meet*, in *Electronic Journal of Comparative Law*, 9(1), 2005, 1-27; LINDA A.J. SENDEN, *Soft law and its implications for institutional balance in the EC*, in *Utrecht Law Review*, 1(2), 2005, 79-99.

²¹ An example of this includes the ethical principles developed by Google, IBM, or Microsoft regarding the use of AI. Self-regulation by private companies raises serious concerns about adequate safeguards in a sector of AI. On one hand, doubts relate to the ability of self-regulation to effectively bind those who set the rules. On the other hand, concerns involve possible situations of vulnerability and imbalance that self-regulation could generate. Regarding the first issue, the capacity of the rules to bind self-regulated parties depends on the strength of the bodies tasked with monitoring compliance, but above all on whether the self-regulating parties perceive it to be in their own interest to comply (so-called *win-win opportunities*). Therefore, incentives for compliance become particularly relevant. As for the second issue, situations of vulnerability and imbalance caused by conflicts of interest, lack of transparency, and limited accountability unfortunately remain a persistent problem that is difficult to address. For a brief analysis of the limitations associated with self-regulation in the field of AI see L. PARONA, *Prospettive europee e internazionali di regolazione dell'intelligenza artificiale tra principi etici, soft law e self-regulation*, in *Rivista della Regolazione dei mercati*, 1, 2020, 70-93. For a critical analysis of de-regulation in the U.S. context see G. RIVLING, *AI Valley: Microsoft, Google, and the Trillion-Dollar Race to Cash In on Artificial Intelligence*, Harper Business, 2025.

²² European Economic and Social Committee, 2015/C 291/05, paragraph 3.4.

²³ European Economic and Social Committee, 2015/C 291/05, paragraphs 1.2 e 2.1. The European Commission started talking about self-regulation and co-regulation around the 2000s with the Better regulation initiative. European

In the governance of AI, the European Union opted for a synergy among all the regulatory instruments described above²⁴. Andrea Simoncini describes the European legal framework on AI as a “*colourful patchwork of different pieces of legislation, employing virtually all regulatory dimensions*”²⁵. The first document to be published on the topic of AI within the European context was the European Parliament’s *Resolution on Civil Law Rules on Robotics*, which made recommendations to the Commission on this specific topic²⁶. In April 2018, twenty-five Member States signed the *Cooperation on Artificial Intelligence* declaration, which set out their intention to collaborate in developing a uniform regulatory framework for AI²⁷. Subsequently, in the same month, the European Commission published a communication presenting its strategy for developing an *AI made in Europe*²⁸. In December 2018, some months later, the European Commission published another communication presenting a *Coordinated Plan on Artificial Intelligence*²⁹. These two documents form the basis of the

Commission, *Communication from the Commission of 25 July 2001 “European governance - A white paper”*, COM(2001) 428 final; European Commission, *Communication from the Commission: European Governance: Better lawmaking*, COM(2002) 275 final; *Impact Assessment*, COM(2002) 276 final; *Consultation document: Towards a reinforced culture of consultation and dialogue - Proposal for general principles and minimum standards for consultation of interested parties* by, COM(2002) 277 final; *Action plan “Simplifying and improving the regulatory environment”*, COM(2002) 278 final, adopted on 5 June 2002. In many cases the European Economic and Social Committee reiterated the advantages, but also the limitations and precise, well-defined objectives associated with the use of self-regulation and co-regulation.

²⁴ For a discussion on the regulation of AI outside of Europe see J.G. CORVALÁN, M.V. CARRO, *Artificial Intelligence Regulation From Outside The European Union: Regulatory Impulses From Other Parts Of The World and a view from Iberoamerica*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, 77-102.

²⁵ A. SIMONCINI, *Verso la regolamentazione della Intelligenza Artificiale*, cited, 415. Own translation.

²⁶ European Parliament, Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). Available at: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html.

²⁷ *Declaration of Cooperation on Artificial Intelligence*, Bruxelles, April 2018. Available at: <https://digital-strategy.ec.europa.eu/en/news/eu-member-states-sign-cooperate-artificial-intelligence>

²⁸ European Commission, *Communication From The Commission To The European Parliament, The European Council, The Council, The European Economic And Social Committee And The Committee Of The Regions. Artificial Intelligence for Europe*, Brussels, 25.4.2018, COM(2018) 237 final. The Strategy’s aim is to promote AI investments and its uptake across Europe. This Strategy represented a fundamental step in the race for AI, where the United States and China were leading the way. The European Union opted for a different approach than the two competitors, moving away both from the US’ laissez-faire approach and the China’s state-centric approach. Some months after the adoption of the Strategy the Joint Research Centre published an interesting report on the European perspective on AI. Joint Research Centre, *Artificial Intelligence: A European Perspective*, 2018. Available at: <https://publications.jrc.ec.europa.eu/repository/handle/JRC113826>.

²⁹ European Commission, *Communication From The Commission To The European Parliament, The European Council, The Council, The European Economic And Social Committee And The Committee Of The Regions. Coordinated Plan On Artificial Intelligence*, Brussels, 7.12.2018, COM(2018) 795 final. With the 2018 Coordinated Plan, the Commission set out the Union’s actions for 2019-2020 in the field of AI. The European Council then presented its conclusions on the Coordinated Plan on Artificial Intelligence. Cfr. Consiglio dell’Unione Europea, *Intelligenza artificiale b) Conclusioni relative al piano coordinato sull'intelligenza artificiale - Adozione*, Bruxelles, 11 February 2019, 6177/19. Available at: <https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/it/pdf>. In 2021 the European Commission revised the Coordinated Plan on AI, presenting it contextually with the Proposal for a Regulation on AI. European Commission, *Annexes to the Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Fostering a European approach to Artificial Intelligence*, Brussels, 21.4.2021, COM(2021) 205 final.

EU's coordinated approach to AI³⁰. In early June 2018, the European Commission launched the initiative *European AI Alliance*³¹ to establish an open policy dialogue on AI. The High-Level Expert Group on Artificial Intelligence (HLEG) was also set up at this time³², which then published the *Ethics Guidelines for Trustworthy AI*³³ and the *Policy and Investment Recommendations for Trustworthy AI*³⁴. In February 2020, the Commission presented a package of policy documents that included a White Paper³⁵, a Report³⁶, and two Communications³⁷. A few months later, in October 2020, the European Parliament presented three resolutions to the Commission with recommendations on the ethical aspects of AI, robotics and related technologies³⁸, a civil liability regime for AI³⁹ and intellectual property

³⁰ The Coordinated Plan reflects the joint commitment of the Commission and the Member States to work together to maximize Europe's potential in AI and to compete globally in this field. It laid the groundwork for cooperation, identified investment areas, and encouraged Member States to develop national AI strategies. Over the following two years, the Plan was revised, and in April 2021, the Commission published an updated version proposing a concrete set of joint actions to establish the EU's global leadership in trustworthy AI. The main objective of the collaboration is to attract public and private investments, which will in turn enable further development.

³¹ The European AI Alliance is a multi-stakeholder online platform seeking broader input on Europe's AI policy.

³² European Commission, *Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions. Building Trust in Human-Centric Artificial Intelligence*, Brussels, 8.4.2019, COM(2019) 168 final. The High-Level Expert Group (HLEG) comprised 52 members from industry, academia, and civil society organisations.

³³ High Level Expert Group, *Ethics Guidelines for Trustworthy AI*, April 2019, which are harmonised ethics guidelines that can be voluntarily followed to develop "trustworthy AI". The guidelines set out seven requirements that an AI system must meet to be considered trustworthy: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) societal and environmental well-being and (7) accountability.

³⁴ High Level Expert Group, *Policy and Investment Recommendations for Trustworthy AI*, June 2019. The Recommendations set out many interesting proposals that then became fundamental pillars of the AI Act. These included the call for a risk-based approach to AI and the importance of taking into account both individual and societal risks. The risk-based approach should be complemented by a precautionary principle-based approach for AI applications that generate unacceptable risks or pose substantial threats of harm.

³⁵ European Commission, *White Paper: On Artificial Intelligence - A European approach to excellence and trust*, Brussels, 19.2.2020, COM(2020) 65 final. The White Paper is the tool the Commission typically uses when it intends to present proposals for Union action in a specific sector. Its purpose is to initiate a discussion with the public, stakeholders, the European Parliament, and the Council with the goal of achieving political consensus for future legislation. The White Paper adopted by the Commission in February 2020 outlines the proposals for action and strategic options to enable the safe and trustworthy development of AI in Europe. Its main goal is to highlight the Union's strengths, identify its weaknesses, and point to future opportunities so that the EU becomes the most attractive, secure, and dynamic data-driven economy in the world. The White Paper introduced the concept of "ecosystems", emphasising that the European Union is trying to foster "ecosystems of excellence" and "ecosystems of trust". The concept of "ecosystem" as it relates to the natural and biological world is not accidental. According to Maria Zanichelli, it is intended to indicate "le interazioni, i flussi e l'equilibrio dinamico che si crea tra gli elementi costitutivi di un sistema digitale, la cui efficacia dipende appunto dall'interconnessione e dall'interoperabilità delle diverse componenti (risorse informatiche, dispositivi, servizi, dati, ma anche gli utenti stessi)". See M. ZANICHELLI, *Ecosistemi, opacità, autonomia: le sfide dell'intelligenza artificiale in alcune proposte recenti della Commissione europea*, 2020, 73.

³⁶ European Commission, *Report From The Commission To The European Parliament, The Council And The European Economic And Social Committee Report On The Safety And Liability Implications Of Artificial Intelligence, The Internet Of Things And Robotics*, Brussels, 19.2.2020 COM(2020) 64 final.

³⁷ European Commission, *Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions. A European strategy for data*, Brussels, 19.2.2020, COM(2020) 66 final; European Commission, *Communication: Shaping Europe's digital future*.

³⁸ European Parliament, *Resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies*, 2020/2012(INL).

³⁹ European Parliament, *Resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence*, (2020/2014(INL).

rights for the development of AI technologies⁴⁰. In April 2021, the European Commission presented the *Proposal for an Artificial Intelligence Act*⁴¹ and negotiations with the other EU institutions began. The AI Act⁴² was finally approved in June 2024, after a long legislative procedure⁴³, and entered into force on the first of August 2024. Although some provisions are already in place, the full implementation of the AI Act is still a long way off⁴⁴.

Following the adoption of numerous soft-law instruments, therefore, the European legislator opted for a comprehensive legal regime to govern AI. Not only did it choose to regulate AI through legally enforceable measures, but it also opted for the most stringent form of legislative instrument available: a regulation⁴⁵. The AI Act has thus general application, is binding in its entirety and directly applicable

⁴⁰ European Parliament, *Resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies*, (2020/2015(INI)).

⁴¹ European Commission, *Proposal for a Regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act)*, Brussels, 21.04.2021, COM (2021) 206 final. For a critical analysis of the Proposal see N.A. SMUHA, *Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea*, in *Philosophy & Technology*, 34, 2021, 91-104; A. MANTELERO, M.S. ESPOSITO, *An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems*, in *Computer Law & Security Review*, 41, 2021, Article 105561, 1-35.

⁴² A reference goes to “actification” of EU law, namely the tendency to give a title to EU legislative acts, such as in the case of the Cybersecurity Act, the Digital Services Act, the Digital Markets Act, the Data Act, Data Governance Act and the AI Act. VAGELIS PAKONSTANTINO, *The actification of EU law: The (long-overdue) move towards eponymous EU legislation*. Published in *European Law Blog*, 26 January 2021. Available at: <https://www.europeanlawblog.eu/pub/the-actification-of-eu-law-the-long-overdue-move-towards-eponymous-eu-legislation/release/1>. See also V. PAKONSTANTINO, P. DE HERT, *The Regulation of Digital Technologies in the EU: The Law-Making Phenomena of 'Actification', 'GDPR Mimesis' and 'EU Law Brutality'*, in *Technology and Regulation Journal*, 50, 2022.

⁴³ On 21 April 2021, the European Commission published a proposal to regulate AI. Public consultations on the Commission’s proposal ended in August 2021. On 6 December 2022, the Council of the EU adopted its common position (“General Approach”) on the AI Act, and on 14 June 2023, the Parliament voted on its negotiating position. On 9 December 2023, the Parliament and the Council reached a provisional agreement on the AI Act. The AI Act was published in the Official Journal of the European Union on 12 July 2024, which formally notifies the new law. The AI Act entered into force on 1 August 2024.

⁴⁴ Prohibitions on certain AI practices and requirements on AI literacy came into effect on 2 February 2025. On 2 August 2025, provisions on notified bodies (Chapter III, Section 4), GPAI models (Chapter V), governance (Chapter VII), confidentiality (Article 78) and penalties (Articles 99 and 100) came into effect. However, providers of GPAI models placed on the market or put into service before 2 August 2025 need to comply with the AI Act by 2 August 2027. 2 August 2025 also represents the deadline for Member States to designate national competent authorities, communicate these to the Commission, and make their contact details publicly available. The remainder of the AI Act will start to apply on 2 August 2026, except Article 6(1), which will start to apply on 2 August 2027. It is noteworthy that the Digital Omnibus on AI Regulation Proposal, presented by the European Commission on 19 November 2025, introduces a revised approach to the timing of applicability of the core requirements for high-risk AI systems. In place of a fixed implementation date, the obligations under the AI Act are now contingent upon the availability of harmonised standards, common specifications, or relevant guidance adopted by the Commission. European Commission, *Proposal for a Regulation Of The European Parliament and of the Council amending Regulations (EU) 2024/1689 and (EU) 2018/1139 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI)*, Brussels, 19.11.2025, COM(2025) 836 final.

⁴⁵ Looking at the wave of digital regulation in the context of the Digital Single Market Strategy, it is easy to see that most of the legislative acts adopted are regulations, preferred over directives. Adopting regulations, rather than directives, helps to reduce the fragmentation of the digital single market. Cfr. M. INGLESE, *Il regolamento sull'intelligenza artificiale come atto per il completamento e il buon funzionamento del mercato interno?*, Quaderni AISDUE - Rivista quadrimestrale, Fascicolo speciale, 2, 2024, 8-9.

in all Member States⁴⁶. Unlike a directive, a regulation does not leave Member States any discretion as to the means and manner of implementation and regulates directly the subject matter it deals with⁴⁷, it does not require the adoption of national implementing measures and is therefore directly effective *vis-à-vis* Member States, public administrations and private operators without the need for further acts. Although the objective was to create a rigid framework applicable to all Member States, the European legislator recognised that a flexible approach was needed to adapt to sudden developments in the AI sector. Without a forward-looking perspective and flexible structure, the legislation would have become obsolete before implementation. To take into account the short-term changes to which AI, or its use, is subject, the European legislator has provided two closely linked regulatory instruments⁴⁸: the Annexes and the Delegated Acts. The advantage of the Annexes is that they can be amended through delegated powers by the Commission under Article 290 TFEU⁴⁹, instead of following the longer ordinary legislative procedure⁵⁰.

The characteristics of regulations ensure the provision of uniform, homogeneous and generally rigid constraints throughout the territory of the Union, thus avoiding different implementation of the legislation between various Member States. However, in the AI Act there is room for *manoeuvre* left to States regarding aspects that do not jeopardise the achievement of the objectives of the AI Act and these specific aspects will therefore be regulated by national legislation⁵¹. In the specific context of law enforcement, for example, the AI Act allows Member States to adopt a national provision authorising the State to use biometric identification systems within the exceptions provided for in Article 5 of the AI Act. Similarly, Member States are called to identify the notifying authorities and

⁴⁶ TFEU, Article 288. M. KELLERBAUER, M. KLAMERT, J. TOMKIN, *Treaty on the Functioning of the European Union*, in M. KELLERBAUER, M. KLAMERT, J. TOMKIN (Eds.), *The EU Treaties and Charter of Fundamental Rights: A Commentary*, Oxford Law Pro, 2024.

⁴⁷ In the Digital Single Market the European legislator opted for the maximum harmonisation, rather than the minimum harmonisation. S. WEATHERILL, *The Fundamental Question of Minimum or Maximum Harmonisation*, in S. GARBEN, I. GOVAERE (Eds.), *The Internal Market 2.0*, Oxford-London, 2020, 261. The European consensus and the so called “internal dimension of the Brussels Effect” increase the externalisation of European rules beyond the borders of Europe.

⁴⁸ These two regulatory instruments are not applied consistently throughout the AI Act. While delegated acts are admissible in relation to high-risk AI systems, the same mechanism does not apply to prohibited AI practices, which are specifically identified in Article 5. However, this diversity is not adequately justified. In fact, even in the case of prohibited AI practices, future developments may lead to the identification of more of them, with the consequence, however, that in that case the amendment will have to be made through the ordinary legislative procedure, rather than through delegated acts. A possible justification is that the political sensitivity of the prohibited AI practices warrants removing this provision from the Commission’s discretionary power, thereby ensuring that any changes must follow the ordinary legislative procedure, where the Council, representing national governments, plays a central role.

⁴⁹ For a critical discussion on the reliance on delegated acts in the AI Act see M. MERLER, *Il ruolo della Commissione Europea nella realizzazione dello spazio digitale europeo: tra costituzionalismo digitale e sovranità digitale*, in *Rivista trimestrale di diritto pubblico*, 4, 2024, 1121-1137.

⁵⁰ AI Act, Article 97.

⁵¹ F. DE CAROLIS, L. TORCHIA, B. MARCHETTI (Eds.), *European Digital Regulation and its impact on Member States*, Springer, 2025. For a first discussion see C. CASONATO, B. MARCHETTI, *Prime osservazioni sulla Proposta di regolamento dell’Unione europea in materia di Intelligenza Artificiale* in *BioLaw Journal*, 3, 2021, 415-437.

the market surveillance authorities for the purpose of Article 70 AI Act⁵². Italy, for example, adopted Law No. 132/2025, which, *inter alia*, designates the competent authorities and defines their respective powers, regulates sectors of national relevance – such as healthcare, the workplace, justice, intellectual property, and criminal sanctions – and delegates to the Government the authority to adopt further implementing provisions across a range of areas. One of the most relevant ones for the purposes of this Thesis is the authority conferred on the Government to adopt a specific regulatory framework governing the use of AI systems in law enforcement. It will be of particular interest to assess whether the Italian legislator will adopt a regulatory framework that, taking due account of concerns identified by academia, European organisations and civil society, will adequately address issues of accuracy, fairness, and transparency.

The AI Act lays down harmonised rules for the placing on the market, the putting into service and the use of AI in the European Union⁵³, and it promotes innovation and the uptake of AI, while ensuring a high protection of health, safety and fundamental rights in the Union⁵⁴. The AI Act is thus the result of a delicate balance between market integration, on the one hand, and protection of fundamental rights and public values, on the other hand⁵⁵. The intention to adopt a regulation that adequately balances these two interests was presented in the European Commission’s White Paper on AI, whose cornerstone was the idea of creating an ecosystem of excellence⁵⁶ and trust⁵⁷ around AI. However,

⁵² With regard to Italy, Law No. 132/2025 identified the Agency for Digital Italy as the notifying authority and the Agency for National Cybersecurity as the market surveillance authority.

⁵³ For these definitions see *Commission Notice – The ‘Blue Guide’ on the implementation of EU product rules 2022*, 2022/C 247/01, Section 2.

⁵⁴ The aim of the European Union’s digital legislation was double: to achieve the digital sovereignty of the EU through the adoption of a regulatory framework that could become a model, and to foster “digital constitutionalism”, protecting EU citizens’ values and rights. See M. MERLER, *Il ruolo della Commissione Europea nella realizzazione dello spazio digitale europeo*, cited.

⁵⁵ The delicate balance between shifting innovation and protecting rights and freedoms was evident since the European Commission’s proposal, whose double aim was to promote investments in the field of AI and promote respect for citizens’ rights. The difficulty in striking a balance between market integration and fundamental rights protection can be seen in many of the legislative acts adopted in the last few years in the digital panorama. Examples are the General Data Protection Regulation, Digital Markets Act, the Digital Services Act.

⁵⁶ Europe’s goal is to foster greater collaboration among cutting-edge research networks and centres to establish a unified hub of expertise that aims to set the global standard for research, innovation, and excellence in AI. To achieve this, the involvement of both the private and public sectors is required. Only through public-private cooperation can the necessary resources to support the entire AI value chain be mobilised. Small and medium-sized enterprises (SMEs) often face greater difficulties in accessing and utilising AI, so adequate incentives must be provided to encourage their involvement. It is also crucial that all public administrations rapidly begin to use AI-based products and services in their operations.

⁵⁷ While acknowledging the positive potential of AI technologies, the Commission is also aware of the risks they pose. It is precisely these unintended consequences that fuel the fears of citizens, businesses and public administrations. The resulting lack of trust is one of the main factors holding back the wider adoption of AI. The White Paper therefore aims to establish a legal framework that protects citizens using AI and provides legal certainty for companies and organisations developing it. To this end, the Commission welcomed the requirements developed by the AI HLEG. AI can only be considered a beneficial strategic technology for citizens, businesses and society as a whole if it is developed in a human-centric, ethical and sustainable way that respects fundamental values and rights. A. BELLOGÍN, O. GRAU, S. LARSSON, G. SCHIMPF, B. SENGUPTA, G. SOLMAZ, *The EU AI Act and the Wager on Trustworthy AI in Communications of the ACM*, 24 November 2022. Available at: <https://cacm.acm.org/research/the-eu-ai-act-and-the-wager-on-trustworthy-ai/>.

despite the European Commission's vision to balance the promotion of innovation with the protection of fundamental rights and values, the underlying structure of the Proposal revealed a predominant focus on industrial safety and market harmonization⁵⁸, rather than on the protection of fundamental rights and public values⁵⁹. It was the Parliament, which passed its amendments in June 2023, that introduced further aspects related to the protection of fundamental rights, such as the reference to a fundamental rights impact assessment⁶⁰. Despite these latter interventions, however, the industrial nature of the AI Act is still evident in the final adopted text⁶¹.

Primary evidence of the fundamental role played by market protection is the choice of the legal basis which supports the AI Act: Article 114 TFEU. In relation to rules on the processing of personal data, Article 16 TFEU which protects the right to data protection complements this legal basis⁶². Article 114

⁵⁸ See A. MANTELERO, *The Artificial Intelligence Act: The European Legislator's Response To The Challenges Of Artificial Intelligence*, cited, 45. There was thus a hiatus between the legal and political discourse, mainly focused on ethical issues and fundamental rights considerations, and the adopted regulatory framework.

⁵⁹ The relationship between regulation and innovation has long been at the center of political and academic debate. This discussion gained renewed prominence at the end of 2024 with the presentation of the *Draghi Report on EU Competitiveness*. Among the key barriers to innovation in Europe, the report identifies regulatory constraints, which are considered particularly burdensome in the technology sector, especially for young and growing companies. MARIO DRAGHI, *The future of European competitiveness*, September 2024. Available at: https://commission.europa.eu/topics/competitiveness/draghi-report_en#paragraph_47059. Responding to concerns about excessive regulatory activity, in November 2025 the European Commission presented a comprehensive digital package aimed at enabling EU businesses to innovate, scale, and reduce administrative costs. The package includes a "Digital omnibus" initiative designed to simplify rules on data governance, cybersecurity, and artificial intelligence, alongside the launch of the Data Union Strategy. A contrasting perspective is offered by Anu Bradford, who challenges the assumption that digital regulation necessarily hampers innovation, arguing instead that regulation and innovation need not be in tension. A. BRADFORD, *The False Choice Between Digital Regulation and Innovation*, in *Northwestern University Law Review*, 119(2), 2024, 377-452.

⁶⁰ Brussels Privacy Hub, *More than 150 university professors from all over Europe and beyond are calling on the European institutions to include a fundamental rights impact assessment in the future regulation on Artificial Intelligence*, 12 September 2023. Available at: <https://brusselsprivacyhub.com/brussels-privacy-hub-and-other-academic-institutions-ask-to-approve-a-fundamental-rights-impact-assessment-in-the-eu-artificial-intelligence-act/>. In addition to a FRIA, Amendment 213 proposed by the European Parliament included general principles for all AI systems, regardless of whether they are high-risk or not, as well as for foundational models: principles of "human intervention and human oversight", "technical robustness and security", "privacy and data governance", "transparency", "diversity, non-discrimination and equity" and "social and environmental welfare". Chapter III of the Council of Europe's *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, deals with "Principles related to activities within the lifecycle of artificial intelligence systems" and sets out the general principles that each Party should apply to its domestic legal system in an appropriate manner. Although somewhat lax, the principles set out in eight articles will apply generally to all types of AI systems.

⁶¹ "Here, in general terms, it is sufficient to note how the industrial policy choices led to a more risk-accepting perspective, different from the more marked risk aversion observed in the GDPR". A. MANTELERO, *The Artificial Intelligence Act*, cited, 46. A critical aspect is also the non-approval of the directive on liability related to the use of AI systems.

⁶² AI Act., Recital 4: "To the extent that this Regulation contains specific rules on the protection of individuals with regard to the processing of personal data concerning restrictions of the use of AI systems for remote biometric identification for the purpose of law enforcement, of the use of AI systems for risk assessments of natural persons for the purpose of law enforcement and of the use of AI systems of biometric categorisation for the purpose of law enforcement, it is appropriate to base this Regulation, in so far as those specific rules are concerned, on Article 16 TFEU". The European Data Protection Board (EDPB) and the European Data Protection Supervisor (EDPS) in Joint Opinion 5/2021 welcomed the reference to this provision, considering it a due reference in cases where the protection of personal data is one of the objectives or essential components of the rules adopted by the European legislator. EDPB-EDPS, *Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, 18 June 2021.

empowers the European Union to enact measures for the harmonisation⁶³ of national rules concerning the establishment and functioning of the internal market⁶⁴. It represents one (*melius* the) cornerstone on which the European Union, and European legislation specifically, has evolved. Article 114 TFEU indeed has represented the fundamental legal basis for the improvement of the Digital Single Market⁶⁵ and the AI Act is no exception⁶⁶.

The legal basis of the AI Act thus underscores its primary nature as a piece of product safety legislation: AI systems are treated as products or components of products that must meet a set of technical and procedural requirements before being placed on the Union market. As such, the AI Act aligns closely with the logic and structure of EU product safety legislation, following the New Legislative Framework approach⁶⁷ and, although regulations adopted under that provision may favour also non-

⁶³ For a definition of the concept “harmonisation” see S. WEATHERILL, *Harmonization in the EU Internal Market*, in *Oxford Encyclopedia of EU Law*, Oxford, 2024, 1.

⁶⁴ Article 114 TFEU authorises the European Parliament and the Council to “adopt the measures for the approximation of the provisions laid down by law, regulation or administrative action in Member States which have as their object the establishment and functioning of the internal market”. The perspective behind the provision is to achieve an advanced level of integration in which EU law serves as the regulatory framework for trade between the Member States. S. MIGLIORINI, *Il ravvicinamento delle legislazioni nella prospettiva del diritto dell’Unione europea e dei suoi effetti sul diritto interno. Le tecniche del ravvicinamento* in *Osservatorio sulle fonti*, 3, 2012, 1.

⁶⁵ On this specific topic see J.-U. FRANCK, G. MONTI, A. DE STREEL, *Options to Strengthen the Control of Acquisitions by Digital Gatekeepers in EU Law*, TILEC Discussion Paper No. DP2021-16, 2021. R. BÖTTNER, H.-J. BLANKE, *Article 114. Harmonisation in the Internal Market*, in R. BÖTTNER, H.-J. BLANKE (Eds.), *Treaty on the Functioning of the European Union - A Commentary*, Volume II, 371-410. Article 114 TFEU is a general legal basis that can only be invoked when the Treaties do not offer a more specific legal basis relevant to a proposed measure. For a discussion on Article 114 TFEU a justified legal basis in line with the EU’s role in safeguarding fundamental rights and signalling a new direction of EU integration see E. LONGO, *Grounding media freedom in the EU: The legal basis of the EMFA*, in *Rivista italiana di informatica e diritto*, 7(1), 2025, 111-124.

⁶⁶ European Commission, *Commission Staff Working Document. Impact Assessment. Accompanying the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts*, Brussels, 21.4.2021, SWD(2021) 84 final, 30 highlights that the AI Act is an «initiative [that] constitutes a core part of the EU single market strategy given that artificial intelligence has already found its way into a vast majority of services and products and will only continue to do so in the future». The choice of article 114 TFEU as the legal basis is motivated as follows “Article 114 TFEU may be used as a legal basis to prevent the occurrence of these obstacles resulting from diverging national laws and approaches how to address the legal uncertainties and gaps in the existing legal frameworks applicable to AI. The present initiative aims to improve the functioning of the internal market by setting harmonized rules on the development, placing on the Union market and the use of products and services making use of the AI technology or provided as stand-alone AI applications. Some Member States are already considering national rules to ensure AI is safe and is developed and used in compliance with fundamental rights obligations. This will likely lead to a further fragmentation of the internal market and increasing legal uncertainty for providers and users on how existing and new rules will apply to AI systems. Furthermore, the Court of Justice has recognised that applying heterogeneous technical requirements could be valid grounds to trigger Article 114 TFEU. The new initiative will aim to address that problem by proposing harmonised technical standards for the implementation of common requirements applicable to the design and development of certain AI systems before they are placed on the market. The initiative will also address the situation after AI systems have been placed on the market by harmonising the way in which ex-post controls are conducted. Based on the above, Article 114 TFEU is the applicable legal basis for the present initiative”. *Ibidem*, 31.

⁶⁷ The New Legislative Framework can be traced back to 2008 with the aim to improve the internal market of industrial products. It consists of three measures: Regulation (EC) 765/2008 which sets out the requirements for accreditation and the market surveillance of products; Decision 768/2008 on a common framework for the marketing of products and Regulation (EU) 2019/1020 on market surveillance and compliance of products. The main objective of the Commission is to bring harmonised legislation on products and related non-legislative acts in line with the Decision 768/2008/EC. Some authors highlighted that the AI Act is even less focused on fundamental rights than other pieces of legislation in the digital

economic considerations, such as those related to fundamental rights protection⁶⁸, its focus remains primarily on the abolition of restrictions on the movement of goods, services, persons and capital⁶⁹, therefore determining the limitations of the AI Act in relation to the protection of fundamental rights and public values⁷⁰.

The nature of a piece of product safety legislation is clearly shown by the scope of the AI Act, that covers the “*placing on the market*”, the “*putting into service*” and the “*use*” of AI systems⁷¹. Furthermore, the AI Act introduces obligations that apply to “providers”⁷² and “deployers”⁷³. A provider, for example, is a developer of an RBI system or of a predictive policing system that markets the system within the Union under its trademark. A deployer might be a law enforcement authority using the RBI system or the predictive policing system in order to maintain public safety. Similarly, a law enforcement authority that manufactures and subsequently uses the system is considered both a provider and a deployer.

It is noteworthy that the AI Act borrows from the GDPR the extraterritorial applicability, which means that the Regulation applies irrespective of whether AI providers or deployers are established or located within the Union or in a third country⁷⁴. The only relevant aspect for the AI Act to apply is that the AI

sphere, such as the Digital Services Act. At the same time, the AI Act limits guarantee for individuals compared to product safety regulation framework. M. ALMADA, N. PETIT, *The Eu AI Act: Between The Rock Of Product Safety And The Hard Place Of Fundamental Rights*, in *Common Market Law Review*, 62, 2025, 85–120.

⁶⁸ In this sense, G. OLIVATO, *The AI Act: regulatory and technical challenges*, PhD Thesis, University of Trento, 20 June 2025, 155. In similar terms, Court of Justice, Judgment of the Court of 5 October 2000, *Federal Republic of Germany v. European Parliament and Council of the European Union*, C-376/98, [2000] ECR I-08419.

⁶⁹ TFEU, Article 26, paragraph 2.

⁷⁰ N.A. SMUHA, K. YEUNG, *The European Union’s AI Act: beyond motherhood and apple pie?*, in N.A. SMUHA (Ed.), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*, Cambridge Law Handbooks, Cambridge University Press, 2025, 234-235. On this specific topic see S. WEATHERILL, *The Limits of Legislative Harmonization Ten Years after Tobacco Advertising: How the Court’s Case Law has become a “Drafting Guide”*, in *German Law Journal*, 12(3), 2011, 827-864.

⁷¹ AI Act, Article 3(9) ‘placing on the market’ means the first making available of an AI system or a general-purpose AI model on the Union market; (10) ‘making available on the market’ means the supply of an AI system or a general-purpose AI model for distribution or use on the Union market in the course of a commercial activity, whether in return for payment or free of charge; (11) ‘putting into service’ means the supply of an AI system for first use directly to the deployer or for own use in the Union for its intended purpose.

⁷² A provider can be a natural or legal person, a public authority, an agency or another body that develops an AI system (or a general-purpose AI model) or that has an AI system (or a general-purpose AI model) developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge. AI Act, Article 3(3).

⁷³ A deployer is a natural or legal person, public authority, agency or other body using an AI system under its authority. The notion of deployer does not cover the person that uses the AI system for a personal non-professional activity.

⁷⁴ On the territorial scope of the AI Act see A.O. GIMÉNEZ, *The Territorial Scope Of Application Of The Artificial Intelligence Act*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 115-129.

system or its output is used within the Union⁷⁵. The extraterritorial application may also exert a positive influence beyond Europe's borders, shaping global standards and practices⁷⁶.

The AI Act's status as product safety legislation is also evident from the introduction of the conformity assessment procedure in the form of the CE marking process⁷⁷. In summary, providers of AI systems must follow a structured process before placing products on the market. First, the AI Act sets out technical requirements (Articles 9-15) designed to reduce risks to safety, health, and fundamental rights. To meet these obligations, providers may rely on harmonised standards or common specifications issued by European standardisation bodies or by the European Commission (see Section 3.4). AI systems must then undergo a conformity assessment to verify compliance with the AI Act's requirements. This assessment is typically conducted by the provider who develops the system. In certain cases, however, it must be carried out by an independent third party – the notified body – following the procedure outlined in Annex VII⁷⁸. Providers issue a declaration of conformity and apply the appropriate marking, clearing the product for market entry. Compliance obligations do not end there, since providers remain responsible for ongoing adherence to the AI Act's requirements after placement on the market, while public authorities ensure enforcement through market surveillance.

While the CE marking process is generally carried out by providers of AI systems, biometric identification systems may require the involvement of third parties in the conformity assessment process. The determining factor is whether the provider has implemented harmonised standards or common specifications to meet the AI Act's requirements: if standards or specifications are not applied, the provider has no discretion and *must* involve a notified body in the conformity assessment. According to Article 43(1) AI Act, in relation to an AI system intended for biometric identification and

⁷⁵ AI Act, Article 2(1).

⁷⁶ On this topic, the most relevant work is A. BRADFORD, *The Brussels Effect: How the European Union Rules the World*, Oxford University Press, 2020. Simply put, the Brussels Effect refers to the tendency of companies to adopt European regulations to maintain access to the EU market. This often leads to a broader impact: not only do companies comply with EU standards (the so-called *de facto* Brussels effect), but also they may adopt similar regulations (so called *de jure* Brussels Effect), amplifying the effect globally. On the relevance of the European consensus as a fundamental aspect for the external dimension of the Brussels Effect see A. BENDIEK, I. STUERZER, *The Brussels Effect, European Regulatory Power and Political Capital: Evidence for Mutually Reinforcing Internal and External Dimensions of the Brussels Effect from the European Digital Policy Debate*, in *Digital Society*, 2, 2023, Article 5. While the Brussels effect is generally acknowledged in relation to data protection legislation, serious doubts arise regarding the same effect in relation to the AI Act because of its differences with data protection legislation. For example, it lacks the provision that allows the European Commission to determine whether a country outside the EU offers an adequate level of protection, as set out in Article 45 of the GDPR.

⁷⁷ Relevant Articles are Articles 8-15, Articles 40 and 41, Article 43, Articles 47 and 48, Articles 70 and 74 to 84. For a detailed description of conformity assessment in the AI Act see A.P. ORTIGOSA, *Conformity Assessment In The Design And Production Of Artificial Intelligence-Based Systems In The Context Of The "New Legislative Framework"*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 449-470.

⁷⁸ If the AI system is a product or a safety component of a product covered by Union harmonisation legislation listed in Annex I, the conformity assessment process for this AI system will be the one envisaged for that product or product safety component in the applicable legislation. The conformity assessment will incorporate verification obligations arising from the AI Act.

intended to be put into service by law enforcement, immigration or asylum authorities, the notified body shall be the market surveillance authority referred to in Article 74(8) or (9). Therefore, the notified body shall be either the competent data supervisory authorities under the GDPR or the LED or any other authority designated pursuant to the conditions laid down in Articles 41 to 44 of the latter⁷⁹.

While the involvement of a notified body for biometric identification systems represents, in my view, a positive safeguard given the high risks those systems may pose to fundamental rights, it is less convincing that other high-risk AI systems used for law enforcement purposes are left to self-assessment by AI providers, given the profound implications such systems have for fundamental rights and freedoms. Indeed, while the verification process must be fully documented and made available to market surveillance authorities and providers are obliged to demonstrate compliance with the AI Act's essential requirements upon a reasoned request, these safeguards seem to appear inadequate to ensure sufficient protection in such a sensitive domain. In any event, it is noteworthy that the European Commission may, through the adoption of delegated acts, subject high-risk AI systems referred to in points 2 to 8 of Annex III to the conformity assessment procedure set out in Annex VII⁸⁰. Looking ahead, it is argued that all AI systems placed on the market, put into service, or deployed for law enforcement purposes should be subject to mandatory third-party conformity assessment, a safeguard that would more effectively guarantee the protection of individuals.

3. The complex legal architecture of the AI Act

Before focusing on the specific domain of law enforcement, it is necessary to have a clear understanding of the structure of the AI Act. The Regulation presents a complex legal architecture, which is the result of the adoption of multiple regulatory approaches which reciprocally influence each other's⁸¹. This legal architecture is considered the main driving force for addressing the challenges that the regulation of AI would inevitably entail in protecting public values and fundamental rights.

⁷⁹ In Italy, the notified body for these AI systems will likely be the Italian Data Protection Authority. See P. STANZIONE, *Segnalazione al Parlamento e al Governo sull'Autorità per l'I.A.*, 25 March 2024. Available at: <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9996508>. Similar considerations have been made regarding the AEPD (the Spanish Data Protection Authority). See ORTIGOSA, *Conformity Assessment In The Design And Production Of Artificial Intelligence-Based Systems In The Context Of The "New Legislative Framework"*, cited, 464.

⁸⁰ AI Act, Article 43(6).

⁸¹ For a detailed discussion on the legal architecture of the AI Act see G. OLIVATO, *The AI Act: regulatory and technical challenges*, cited, Chapters 2 and 3.

3.1. The horizontal approach

One of the first decisions that policymakers have to take when regulating AI is whether to adopt a horizontal or vertical approach⁸². A horizontal approach introduces general principles and uniform requirements that apply across all sectors regardless of the specific context in which the AI system is deployed. In contrast, a vertical approach focuses on specific AI applications and regulates them through tailored obligations and safeguards, considering their unique characteristics and sector-specific risks⁸³. The vertical approach, despite its advantages, carries a significant risk of over-fragmentation, one that could prove damaging to market stability and broader commercial interests. This could take the form of legal inconsistencies and uncertainties, higher compliance costs, and non-uniform standards and principles. These issues would be particularly detrimental to cross-sectoral providers and would affect the protection of fundamental rights, lowering protection and safeguards across sectors.

In an overall assessment, the European legislator considered horizontal regulation to be more appropriate to avoid the fragmentation of the Digital Single Market and to ensure the harmonisation of provisions on AI⁸⁴. Therefore, the AI Act applies horizontally, with limited exceptions⁸⁵, categorizing AI applications into four categories of risk, plus a single body of rules for general-purpose AI models (GPAI)⁸⁶, each of which is governed by specific requirements, obligations and supplemented by a few general rules⁸⁷. The choice of a single piece of legislation inevitably added

⁸² M. O'SHAUGHNESSY, M. SHEEHAN, *Lessons From the World's Two Experiments in AI Governance*, 14 February 2023. Available at: <https://carnegieendowment.org/posts/2023/02/lessons-from-the-worlds-two-experiments-in-ai-governance?lang=en>.

⁸³ For an interesting discussion about the rules applicable to AI see G. FINOCCHIARO, *Intelligenza Artificiale, quali regole?*, il Mulino, 2024.

⁸⁴ In the Impact Assessment published in April 2021 the European Commission identified multiple options: Option 1: EU Voluntary labelling scheme; Option 2: Ad hoc sectoral approach; Option 3: Horizontal risk-based act on AI; Option 3+: Horizontal EU legislative instrument establishing mandatory requirements for high-risk AI applications plus voluntary Codes of Conduct for non-high risk applications; Option 4: Horizontal EU legislative instrument establishing mandatory requirements for all AI applications, irrespective of the risk they pose. The impact assessment provides also an analysis of every single option. European Commission, *Commission Staff Working Document Impact Assessment - Accompanying the Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts*, Brussels, 21.4.202, SWD(2021) 84 final, COM(2021) 206 final. In relation to the horizontal approach (Option 3 and Option 3+) it is stressed that such «rules would be uniform and harmonised across all Member States which would address the problems of legal uncertainty and market fragmentation and help citizens and companies build trust in the AI technology placed on the Union market».

⁸⁵ AI Act, article 2(3). It does not apply to AI systems for national security, defence and military purposes, since they remain within the competences of the Member States, and for the sole purpose of scientific research.

⁸⁶ For a detailed discussion of requirements and obligations related to GPAI models see J.A. CASTILLO PARRILLA, *General-Purpose Artificial Intelligence, Foundational Models (And "GPT Chat") in the Artificial Intelligence Act* in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 801-824.

⁸⁷ Chapter 1 sets out the general provisions that apply to AI systems, regardless of their risk category. Some authors consider this approach to introduce a complex system, which, rather than simplifying, «threatens to make the landscape of industrial exploitation of AI tortuous and open to litigation». A. MANTELERO, *The Artificial Intelligence Act: The European Legislator's Response To The Challenges Of Artificial Intelligence*, cited, 48.

complexity to the governance of AI, as the regulatory requirements the AI Act contains must adapt appropriately to the different types, sectors and uses of AI.

Despite this choice generally seems to be appropriate, the horizontal approach may underplay the higher stakes involved in some specific contexts, such as in law enforcement. In this domain, a vertical regulatory approach built upon a horizontal foundation would, in my assessment, be preferable. While the horizontal framework establishes general principles and rules, sector-specific rules for law enforcement might address unique ethical, legal, and societal concerns that the use of AI for law enforcement purposes entails. The AI Act itself appears to recognise this necessity. Although no sector-specific legislation has been adopted, the distinctive characteristics of the law enforcement sector have led the European legislator to establish an intricate framework of rules, prohibitions, and exceptions, effectively creating a form of special regulation for AI systems in this domain. This tendency is not new. In the field of data protection legislation, for example, the specific characteristics of law enforcement prompted the European legislator to adopt the already mentioned LED, which complements the GDPR by addressing the particularities of processing personal data for law enforcement purposes.

An example of vertical (or sectoral) regulation is the one concerning face recognition technology (hereinafter, FRT). Adopting a global perspective, in the last five years, many countries have considered to regulate FRT with specific regulation. In 2020, for instance, the State of Washington enacted legislation requiring state and local agencies, including law enforcement authorities, that use, or intend to use, FRT to comply with specific reporting and deployment requirements⁸⁸. Similarly, the State of Colorado⁸⁹ and the State of Maryland⁹⁰ enacted legislation imposing strict regulations on the use of FRT by government entities and law enforcement authorities. Also other States have opted to temporarily or otherwise limit law enforcement use of FRT⁹¹. In 2019, San Francisco became the first U.S. city to ban government use of FRT, followed by other municipalities including Boston, Portland, Oregon⁹². In general, the adoption of specific rules on FRT is considered necessary by United Nations

⁸⁸ State of Washington, S6280, 2020. Available at: <https://lawfilesexternal.wa.gov/biennium/2019-20/Pdf/Bills/Senate%20Bills/6280.pdf>. See also: <https://app.leg.wa.gov/RCW/default.aspx?cite=43.386&full=true>.

⁸⁹ State of Colorado, S113, 2022. Available at: https://leg.colorado.gov/sites/default/files/documents/2022A/bills/2022a_113_enr.pdf. The bill of the State of Washington and the State of Colorado require an accountability report, data management, security protocols, training procedures and testing, for government entities to use FRT. Additionally, entities must obtain a warrant or court order to use the technology to conduct ongoing surveillance, real-time identification or tracking.

⁹⁰ State of Maryland, SB182, 2024. Available at: <https://mgaleg.maryland.gov/mgawebsite/Search/FullText>.

⁹¹ For a detailed analysis of the status of state laws on FRT see <https://www.techpolicy.press/status-of-state-laws-on-facial-recognition-surveillance-continued-progress-and-smart-innovations/>. For a critical analysis of legislation on FRT in the United States see B. ALLYN, *With no federal facial recognition law, states rush to fill void*, 28 August 2025. Available at: <https://www.npr.org/2025/08/28/nx-s1-5519756/biometrics-facial-recognition-laws-privacy>.

⁹² For an in-depth representation of different U.S. nationwide and citywide regulations about biometric recognition, see <https://www.banfacialrecognition.com/map/>. The different layers of biometric recognition related laws in the USA is

experts, who have called for an immediate moratorium on the sale, transfer, and use of such tools because of the unacceptable risks they pose to fundamental rights⁹³. Similarly, the European Data Protection Board (EDPB) and the European Data Protection Supervisor (EDPS) jointly called for a general prohibition on the use of FRT in their Opinion No. 5/2021. The EDPB has further addressed the fundamental rights implications of such technology in its Guidelines No. 5/2022, recognising that its deployment is liable to affect a broad range of fundamental rights and freedoms. In 2021, the Council of Europe called for a moratorium on the use of live FRT in publicly accessible spaces (such as streets, shopping centres, and schools) pending the establishment of adequate legal safeguards⁹⁴. In the same year, the European Parliament advocated for strict limitations on FRT⁹⁵, while the European Commission, in its impact assessment, took the view that remote biometric identification systems warranted an outright ban⁹⁶. More broadly, the use of face recognition for law enforcement purposes remains among the most contested applications of AI⁹⁷.

It is unsurprising thus that although the AI Act was conceived as a horizontal instrument of general application across sectors, it nonetheless established an elaborate, special regulatory framework for AI systems used in law enforcement. Similarly, the recently enacted Italian Law No. 132/2025 empowers the Government to develop a separate regulatory framework for the deployment of AI systems in such domain⁹⁸, specifically recognising the peculiarities of this sector. Inevitably, the normative significance of this delegation will ultimately depend on how the implementing legislation is worded.

discussed in M. FIDLER, J. HURWITZ, *An Overview of Facial Recognition Technology Regulation in the United States*, in R. MATULIONYTE, M. ZALNIERIUTE (Eds.), *The Cambridge Handbook of Facial Recognition in the Modern State*, Cambridge Law Handbooks, Cambridge University Press, 2024, 214-227.

⁹³ In 2019, it was David Kaye, the Special Rapporteur on freedom of opinion and expression. *UN expert calls for immediate moratorium on the sale, transfer and use of surveillance tools*, 25 June 2019. The Report is available at: <https://www.ohchr.org/en/press-releases/2019/06/un-expert-calls-immediate-moratorium-sale-transfer-and-use-surveillance>. In 2021, the UN High Commissioner for Human Rights, Michelle Bachelet, called for such action. *Urgent action needed over artificial intelligence risks to human rights*, 15 September 2021. Available at: <https://news.un.org/en/story/2021/09/1099972>. M. ZALNIERIUTE, *Facial recognition surveillance and public space: protecting protest movements*, in *International Review of Law, Computers & Technology*, 39(1), 2025, 116–135; M. O’FLAHERTY, *Opinions, Facial Recognition Technology and Fundamental Rights*, in *European Data Protection Law Review*, 6(2), 2020, 170.

⁹⁴ Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data (Convention 108), *Guidelines on Facial Recognition*, 2021. Available at: <https://edoc.coe.int/en/artificial-intelligence/9753-guidelines-on-facial-recognition.html>.

⁹⁵ T. MADIEGA, H. MILDEBRATH, *Regulating facial recognition in the EU*, *European Parliament*, cited, 2021.

⁹⁶ In the impact assessment presented by the European Commission remote biometric identifications systems were considered a “very specific and urgent case that requires regulatory intervention”. European Commission, *Commission Staff Working Document Impact Assessment*, cited, 45.

⁹⁷ FRA, *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, 2019. European Digital Rights (EDRi), *How to fight Biometric Mass Surveillance after the AI Act: A legal and practical guide*, March 2024. Available at: <https://edri.org/our-work/how-to-fight-biometric-mass-surveillance-after-the-ai-act-a-legal-and-practical-guide/>.

⁹⁸ *Disposizioni e deleghe al Governo in materia di intelligenza Artificiale*, article 24 letter h). Text approved by the Italian Senate available at: <https://www.senato.it/leggi-c-documenti/disegni-di-legge/scheda-ddl?tab=testiEmendamenti&did=59313>.

3.2. The risk-based approach

The AI Act follows a risk-based approach, categorising AI systems based on the level of risk they pose⁹⁹. Risk is commonly defined as “*the combination of the probability of an occurrence of harm and the severity of that harm*”¹⁰⁰. Therefore, while *risk* is prospective, future-oriented, uncertain, and denoting the possibility or likelihood of damage, *harm* is retrospective, referring to concrete, realized negative impacts that have already occurred. In AI governance, however, these categories often get collapsed into “risks”¹⁰¹. The concept of “risk” lies at the core of “risk regulation”¹⁰², that is all regulatory interventions aimed at anticipating potential harms to individuals or society arising from certain technologies, activities, or products¹⁰³. The overarching goal of such regulatory response is to effectively assess, manage and mitigate potential harms *before* they materialize. In the case of AI, while it is widely acknowledged that this family of technologies can pose risks to both individuals, groups of individuals and society, the nature and severity of them remain only partially understood. For this reason, lawmakers around the world have converged around the model of risk regulation to govern AI¹⁰⁴, as opposed to *ex post* liability regime¹⁰⁵.

⁹⁹ The risk-based approach is the starting point of most of the legislative interventions in the digital realm. It is the case of the GDPR, the Cybersecurity Act, the Network Information Security Directive (NIS2), the DORA Directive.

¹⁰⁰ AI Act, Article 3(2). This definition is reproduced by ISO/IEC GUIDE 51:2014, *Safety aspects — Guidelines for their inclusion in standards*, 3. A relevant definition of risk can be found in ISO 31000:2018 Risk management — Guidelines, Edition 2, 2018.

¹⁰¹ M. KAMINSKY, *Regulating the Risks of AI*, in *Boston University Law Review*, 103, 2023, 1347;1367. Policymakers and companies seem to prefer to talk about risks because it emphasizes management and mitigation before harms occur. Furthermore, many frameworks (like the EU AI Act, ISO standards, or OECD principles) are structured around risk management rather than retrospective harm analysis. AI harms are often diffuse, indirect, and systemic. Indeed, it is hard to point to a single event as “the harm”, so they’re preferably translated into ongoing risks. Industry actors are already embedded in risk assessment frameworks (compliance, liability, audits), so framing AI harms as “risks” fits existing governance tools. The trade-off is that by framing everything as “risk,” actual experienced harms can be downplayed and abstracted.

¹⁰² Among others, see J. BLACK, *The role of risk in regulatory processes*, in R. BALDWIN, M. CAVE, M. LODGE (Eds.), *The Oxford Handbook of Regulation*, Oxford University Press, 2010, 302-348.

¹⁰³ Examples of risk regulation are Product Safety legislation (Directive (2001/95/EC), Food Safety legislation (Regulation (EC) No 178/2002), Medical Devices legislation (Regulation (EU) 2017/745), Pharmaceuticals legislation (Directive 2001/83/EC and Regulation (EC) No 726/2004). The *European Journal of Risk Regulation* is a journal devoted entirely to examining regulatory responses to both traditional and emerging risk phenomena.

¹⁰⁴ One of the most interesting works on the topic is KAMINSKY, *Regulating the Risks of AI*, cited, and M. KAMINSKY, *The Developing Law of AI: A Turn to Risk Regulation*, in *The Digital Social Contract: A Lawfare Paper Series*, April 2023. M.U. SCHERER, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, in *Harvard Journal of Law & Technology*, 29(2), 2016, 354-400. Apart from the AI Act, other documents are based on the risk regulation approach. See UK Government, *The Bletchley Declaration by Countries Attending the AI Safety Summit*, 1-2 November 2023. Available at <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration>.

¹⁰⁵ AI systems are complex resulting in technically and legally opaque systems. Moreover, AI harms might not always be foreseeable, therefore it can be particularly hard to determine causality *ex post*. In addition, *ex ante* regulation is well suited to AI systems since it “*has the potential to influence or even dictate system design, addressing harms on a collective level*”

The risk regulation approach to AI governance presents both advantages and limitations¹⁰⁶. On one hand, it offers a pragmatic and flexible framework that calibrates regulatory intervention according to the potential harm posed by different AI systems, avoiding blanket restrictions that could stifle innovation and offering interesting tools, such as precautionary tactics¹⁰⁷ and post-market measures¹⁰⁸. On the other hand, this approach faces significant challenges: defining risk categories can be inherently subjective, enforcement mechanisms may prove inadequate, and the framework's reliance on *ex ante* assessments may fail to capture emergent risks that only become apparent through deployment and real-world use. Moreover, risk regulation is most effective when addressing quantifiable risks¹⁰⁹. In the context of AI, however, risks are often difficult to quantify¹¹⁰ and, in some cases, entirely unforeseeable (so called “unknown unknowns”)¹¹¹. As far as the former aspect is concerned, risks for fundamental rights, in fact, are generally not quantitative, but qualitative (e.g. how a decision affects someone's dignity or autonomy); they are context-sensitive and based on subjective interpretation; therefore, they are harder to define, predict, or measure objectively and this makes it challenging to apply traditional risk assessment tools to such risks¹¹². As far as the second aspect is concerned, the adaptability of AI¹¹³ and of its uses make it difficult to foresee all possible risks that the AI system could entail. Another limitation is that risk regulation typically focuses on collective harm, that is harm affecting groups of individuals who share common characteristics. As a result, it tends to overlook individual harm, that is the one that affects specific individuals (e.g. an individual is arrested due to misidentification by an RBI system), and societal harm, the one that affects the society as a whole¹¹⁴.

rather than waiting for individuals to invoke their rights”. KAMINSKY, *The Developing Law of AI: A Turn to Risk Regulation*, cited, 1367.

¹⁰⁶ Kaminski, for example, critically examines the application of risk regulation to AI: “*There are good reasons to use at least some risk regulation to govern AI systems; but we must be aware of its blind spots, quirks, and shortcomings as a set of regulatory tools*”. KAMINSKY, *The Developing Law of AI: A Turn to Risk Regulation*, cited, 1365.

¹⁰⁷ Which include legal bans, licensing, and regulatory sandboxing.

¹⁰⁸ It includes revocable licenses, registration coupled with post-market monitoring, cyclical evaluation for compliance with performance metrics, and failsafe mode.

¹⁰⁹ KAMINSKY, *Regulating the Risks of AI*, cited, 1373

¹¹⁰ For instance, it is the case of the erosion of anonymity in public spaces due to the use of AI-powered surveillance cameras or the spread of disinformation due to fake news or polarized news.

¹¹¹ It is the case of risks that do not arise during the design or development phase of the AI system, but they arise after the system is put on the market or put into service. It is, for example, the case of Large Languages Models which have versatile and unpredictable applications once they are put into service and used.

¹¹² Some authors proposed to apply the risk assessment in climate change – the Intergovernmental Panel on Climate Change (IPCC) model – to AI systems on the basis of the fact that climate change risk and AI risk share some similarities. C. NOVELLI, F. CASOLARI, A. ROTOLO, M. TADDEO, L. FLORIDI, *Taking AI risks seriously: a new assessment model for the AI Act*, in *AI & SOCIETY*, 2023. Available at: <https://doi.org/10.1007/s00146-023-01723-z>. See also J. BLACK, R. BALDWIN, *Really Responsive Risk-Based Regulation*, in *Law and Police*, 32(2), 2010, 181-231. DOI: 10.1111/j.1467-9930.2010.00318.x.

¹¹³ AI Act, Article 3(1) highlights that adaptability is one of the main characteristics of an AI system.

¹¹⁴ AI does not only create potential harm to individuals and to groups of individuals, but it also creates risk for the society. It is the case of increased surveillance and negative implications for democracy due to the spread of disinformation. N.A. SMUHA, *Beyond the individual: governing AI's societal harm*, in *Internet Policy Review*, 10(3), 2021, 1-32. Available at: <https://doi.org/10.14763/2021.3.1574>.

In addition, adopting a risk-based regulatory model inherently means accepting that AI systems may be placed on the market, put into service, or used even when they carry certain risks (residual risks) that cannot be fully eliminated¹¹⁵. Indeed, the obligations and requirements set out in the AI Act are primarily designed to mitigate and manage these risks to an acceptable level, rather than to achieve their total eradication. When regulation centres primarily on *ex ante* obligations – such as preventive risk management rules and a *by design* approach – it inherently calls for a robust liability framework to address the residual risks that persist even after those preventive measures are in place. Specifically, it should account for situations in which harm results from incorrect, inadequate, or poorly implemented risk management. Without such a complementary system of liability, the regulation may fall short in providing sufficient accountability and redress for affected parties¹¹⁶. In fact, even though residual risks cannot be excluded, *ex post* additional measures, such as compensation, should be required¹¹⁷. Certainly, the AI Act introduces an entire Chapter XII dedicated to penalties, which are to be established by Member States in accordance with the criteria outlined in Articles 99, 100, and 101. However, these penalties are primarily aimed at ensuring compliance by providers and deployers of AI systems and they do not provide direct recourse or compensation for individuals who may have been harmed because of the use of AI systems. Moreover, the remedies outlined in Section 4 of Chapter IX are notably limited. While they include the right to lodge a complaint with a market surveillance authority, the right to an explanation of individual decision-making, and the protections afforded by Directive (EU) 2019/1937 on whistleblower protection, these provisions fall short of establishing comprehensive individual rights or meaningful redress in cases of harm caused by AI systems. A core feature of risk regulation is its foundation on the principle of proportionality¹¹⁸. Aligned with Article 5(4) TEU, the AI Act adopts a differentiated regulatory framework based on the potential

¹¹⁵ “[...] risk regulation typically assumes a technology will be adopted despite its harm”. KAMINSKY, *Regulating the Risks of AI*, cited, 1352. The high-risk requirement related to the introduction of a risk management system precisely stresses the fact that risks still remains and that it is the provider of the AI system that shall identify the appropriate measures to mitigate those risks.

¹¹⁶ MANTELERO, *The Artificial Intelligence Act: The European Legislator’s Response To The Challenges Of Artificial Intelligence*, cited, 47. «Developing an *ex ante* protection model, focused on risk analysis, without then developing an adequate framework for the residual hypotheses in which poor or deficient risk management causes damage, ends up undermining the overall impact of the regulatory intervention derived from the AIA, which is thus an unfinished work in a broad view of the regulation of AI».

¹¹⁷ On 28 September 2022, the European Commission published a proposal for an Artificial Intelligence Liability Directive (“AI Liability Directive”), intended to address claims for harm caused by AI systems, by adapting non-contractual civil liability rules to the specific challenges posed by AI. However, in its 2025 Work Programme, the European Commission effectively scrapped the AI Liability Directive (“AILD”). Cfr. K. ZENNER, *An AI Liability Regulation would complete the EU’s AI strategy*, 25 February 2025. Available at: <https://www.ceps.eu/an-ai-liability-regulation-would-complete-the-eus-ai-strategy/>

¹¹⁸ TEU, Article 5(4): “Under the principle of proportionality, the content and form of Union action shall not exceed what is necessary to achieve the objectives of the Treaties”. Consolidated version of the Treaty on the Functioning of the European Union – Protocol (No 2) on the application of the principles of subsidiarity and proportionality, OJ C 115, 9.5.2008.

impact of AI systems on individuals' health, safety, and fundamental rights¹¹⁹. This approach introduces tailored obligations and requirements depending on the specific AI system and its intended use. Such a structure enables regulators to focus their efforts and resources on the applications that pose the highest risks, while avoiding unnecessary overregulation where it is not warranted¹²⁰. Alongside proportionality, the precautionary principle also plays a key role, underpinning the prohibition of AI practices deemed incompatible with the values of the European Union¹²¹.

Within the AI Act, the risk-based approach is applied with certain distinct features¹²². Initially, the legislator, based on a prior assessment, identified four categories of risk, determined by the intended purpose and context of use of the AI system. It is the Regulation itself that explicitly identifies these categories and distinguishes between AI systems that pose an unacceptable risk, a high risk, a limited or minimal risk. Besides, the AI Act introduces harmonised rules for the placing on the market of general-purpose AI models. It is possible that, over time, adjustments to the existing risk categories may become necessary, whether within a given category or across categories, because of technological developments or shifting policy priorities. However, mechanisms for modifying risk-categories differ: while the list of prohibited AI practices can only be amended through the ordinary legislative procedure, which requires a full involvement of the European Parliament and the Council, the list of use-cases within the domains in Annex III may be updated more flexibly by the European Commission through delegated acts¹²³. It is noteworthy to highlight that, while use-cases within each domain can be updated by the European Commission through delegated acts, the domains themselves cannot and require the ordinary legislative procedure to be followed¹²⁴.

In the first instance, therefore, it is the European legislator who determines the relevant categories of risk *ex ante*, thereby setting the regulatory framework within which AI systems are classified. Scholars have observed that the AI Act's risk-based approach is essentially top-down: it is the legislator who

¹¹⁹ Some authors highlighted that the risk-based approach is inherently complex due to “the difficulty of providing a precise definition of high-risk systems and the evolving uses of AI”. MANTELERO, *The Artificial Intelligence Act: The European Legislator's Response To The Challenges Of Artificial Intelligence*, cited, 47. However, the rationale of the risk-based approach was not to “make it easier for operators to know from the out set whether or not they are subject to the new rules”, rather to avoid overregulation in areas where it may not be necessary.

¹²⁰ See J. BLACK, *Risk-based regulation: choices, practices and lessons learned*, in *Risk and Regulatory Policy. Improving the Governance of Risk*, OECD, 2010, 186. Available at: <https://eprints.lse.ac.uk/36387/>.

¹²¹ OLIVATO, *The AI Act: regulatory and technical challenges*, cited, 112.

¹²² T. MAHLER, *Between Risk Management and Proportionality: The Risk-Based Approach in the EU's Artificial Intelligence Act Proposal*, in *Nordic Yearbook for ICT Law*, 2022, 251; G. DE GREGORIO, P. DUNN, *The European Risk-based Approaches: Connecting Constitutional Dots in the Digital Age*, in *Common Market Law Review*, 59, 2022, 473-475.

¹²³ AI Act, Article 7. Despite this difference might reflect a balance between democratic oversight for prohibitions of unacceptable AI practices and regulatory agility for adapting the high-risk category to emerging developments, this choice has been criticised since it implies the amendment of a core part of the AI Act. MANTELERO, *The Artificial Intelligence Act: The European Legislator's Response To The Challenges Of Artificial Intelligence*, cited, 49.

¹²⁴ AI Act, Article 7.

pre-defines the levels of risk and assigns corresponding legal obligations to each category¹²⁵. However, this reliance on predefined risk levels, determined by the purpose and context of use of an AI system, may pose challenges for the protection of fundamental rights. Certain AI systems may raise genuine fundamental rights concerns and yet fall within the limited-risk or even minimal-risk categories, a classification that risks underestimating their actual harm potential. Place-based predictive policing is one such example.

In addition, determining the appropriate risk level for AI systems is not straightforward¹²⁶. This is evident in the case of person-based predictive policing, where a distinction is made between systems that assess or predict the likelihood of an individual committing a criminal offence based *solely* on profiling or on the evaluation of personality traits and characteristics, and those that do not rely exclusively on such profiling or evaluation. While these two types of systems fall into different risk categories under the AI Act, in practice, it is not always clear whether a given system fits into one category or the other, making classification ambiguous and potentially inconsistent.

To limit these issues, the top-down approach is complemented by bottom-up assessments carried out by providers and deployers of AI systems, respectively under Article 9 and Article 27. Therefore, while the legislator determines the overall categories of risk and corresponding obligations, it is up to individual actors to conduct system-specific evaluations in order to identify, document, and mitigate the concrete risks that may arise in practice¹²⁷. In this way, the AI Act combines regulatory classification with contextualised self-assessment to create a multi-layered risk governance framework.

At the lowest end of the spectrum, systems considered to present limited or minimal risk are subject to minimal obligations, preserving space for innovation and reducing unnecessary regulatory burdens¹²⁸. In relation to limited risk AI systems, the AI Act identifies transparency obligations under

¹²⁵ For a discussion on this aspect see OLIVATO, *The AI Act: regulatory and technical challenges*, cited, 104.

¹²⁶ ALMADA, PETIT, *The Eu AI Act: Between The Rock Of Product Safety And The Hard Place Of Fundamental Rights*, cited, 11.

¹²⁷ Through the risk management process under Article 9 or the Fundamental Rights Impact Assessment (FRIA) under Article 27. Similarly, in the USA, the National Institute for Standard and Technology published the *Risk Management Framework* (AI RMF 1.0) and the *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. They are applied on a voluntary basis, and they are not specifically focused on fundamental rights. In the specific context of AI, the favour for employing a risk-based methodology to guide the assessment of risks associated with AI technologies, has been highlighted in multiple documents. See OECD, *Recommendation of the Council on Artificial Intelligence*, May 2024. Available at: <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>. G20 Ministerial Statement on Trade and Digital Economy, 9 June 2019, available at <https://wp.oecd.ai/app/uploads/2021/06/G20-AI-Principles.pdf>. High Level Expert Group On Artificial Intelligence, *Ethics guidelines for trustworthy AI*, cited.

¹²⁸ Limited or minimal-risk AI systems represent the majority. The European Commission estimated that 90% of AI systems and two thirds of AI systems that will be used in the public sector will not be classified as high risk. European Commission, *Study to Support an Impact Assessment of Regulatory Requirements for Artificial Intelligence in Europe*, Final Report (D5), April 2021, 143. Available at: <https://op.europa.eu/en/publication-detail/-/publication/55538b70-a638-11eb-9585-01aa75ed71a1/language-en>. L. TANGI ET AL., *AI Watch European landscape on the use of Artificial Intelligence by the Public Sector*, JRC Science For Policy Report, European Union. 2022, 58. For a detailed discussion on the regulation of

Article 50 AI Act. For minimal risk AI systems, the AI Act does not provide for any regulatory obligation, but providers and deployers may adhere to voluntary Codes of Conduct¹²⁹.

In contrast, at the highest end of the spectrum, lies Article 5, which deals with prohibited AI practices, where public authorities are generally involved. Practices included in this provision are considered particularly harmful by the European legislator since they contradict European values enshrined in Article 2 TUE – respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities – and rights and freedoms incorporated in the CFREU¹³⁰. In relation to these prohibitions, the AI Act adopts a precautionary approach¹³¹, denying the presence on the market or the use of AI systems whenever the risks stemming from their use are deemed too high for risk-mitigating interventions. Article 5 is one of the provisions that underwent the most changes during the negotiation process¹³², given that the decision to prohibit certain AI practices is inherently politically sensitive¹³³. In February 2025, to ensure consistent, effective and uniform application among Member States, the European Commission published the *Guidelines on Prohibited Artificial Intelligence Practices Established by Regulation (EU)*

such systems in the AI Act see L. COTINO HUESO, *Codes of conduct, seals or certifications for artificial intelligence systems that are not high risk (article 95 of the AI Act)*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 825-839. Just because the AI Act does not introduce stringent requirements or obligations does not mean that AI systems are completely unregulated. Indeed, if AI systems process personal data, data protection legislation applies. Furthermore, Regulation (EU) 2023/988 of the European Parliament and of the Council of 10 May 2023 on general product safety also applies.

¹²⁹ AI Act, Article 95.

¹³⁰ Among the most relevant rights are: the right to private and family life (Article 7), data protection (Article 8), freedom of thought, conscience and religion (Article 10), right to freedom of expression and information (Article 11), freedom of assembly and of association (Article 12), right to equality (Article 20), right to non discrimination (Article 21), right to an effective remedy and fair trial (Article 47), presumption of innocence and right of defence (Article 48). Similarly, the Council of Europe's *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* considers unacceptable those AI systems which are deemed «incompatible with the respect for human rights, the functioning of democracy or the rule of law». In line with the Convention, the UN AI Resolution calls to refrain or cease the use of AI systems which are “impossible to operate in compliance with international human rights law or that pose undue risks to the enjoyment of human rights, especially of those who are in vulnerable situations». The prohibition of such practices is in line with the UNESCO *Recommendation on the Ethics of Artificial Intelligence* adopted in 2021 and with the 2021 report of the United Nations High Commissioner for Human Rights Michelle Bachelet.

¹³¹ On the role of precautionary principle in EU regulation, see K. GARNETT, D.J. PARSONS, *Multi-Case Review of the Application of the Precautionary Principle in European Union Law and Case Law*, in *Risk Analysis*, 37(3), 2017, 502-516.

¹³² The final text of the AI Act prohibits practices that involve AI systems: a) deploying subliminal, manipulative, or deceptive techniques to distort behaviour and impair informed decision-making, causing significant harm; b) exploiting vulnerabilities related to age, disability, or socio-economic circumstances to distort behaviour, causing significant harm; c) biometric categorisation systems inferring sensitive attributes (race, political opinions, trade union membership, religious or philosophical beliefs, sex life, or sexual orientation); d) assessing the risk of an individual committing criminal offenses solely based on profiling or personality traits, except when used to augment human assessments based on objective, verifiable facts directly linked to criminal activity; e) compiling facial recognition databases by untargeted scraping of facial images from the internet or CCTV footage; f) inferring emotions in workplaces or educational institutions, except for medical or safety reasons; g) evaluating or classifying individuals or groups based on social behaviour or personal traits, causing detrimental or unfavourable treatment of those people; h) ‘real-time’ remote biometric identification (RBI) in publicly accessible spaces for law enforcement, with some exceptions.

¹³³ In the end, prohibited AI practices are the result of lobbying efforts from diverging interests and affected stakeholders. Therefore, the risk is to leave gaps in the AI systems’ coverage.

2024/1689¹³⁴, providing guidelines for the practical implementation of Article 5 of the AI Act¹³⁵. Although those Guidelines are non-binding¹³⁶, they represent a useful tool for public and private subjects who are providers, deployers or are otherwise involved in the AI value chain in order to correctly apply the provision.

Between these extremes lies the category of high-risk AI systems¹³⁷, which may be placed on the market, put into service and used only if they comply with a set of strict requirements¹³⁸. The AI Act identifies two categories of high-risk AI systems. A first category, comprehend AI systems that are intended to be used as a safety component of a product, or are themselves a product, covered by the Union harmonisation legislation listed in Annex I of the AI Act¹³⁹. A second major category of high-risk AI systems, as defined in the AI Act, comprises stand-alone AI applications deployed in sensitive and high-impact domains (points 2-8), that are: critical infrastructure; education and vocational training; education, workers management and access to self-employment; access to and enjoyment of essential private services and essential public services and benefits; law enforcement; migration, asylum and border control management; administration of justice and democratic processes. As is evident from the list above, the European legislator chose to classify as high-risk those AI systems primarily employed in the exercise of public functions and the delivery of public services¹⁴⁰. Moreover, point 1) refers to biometrics as a horizontal category that applies across all domains.

The determining element of Annex III is the “intended purpose”¹⁴¹ for which the AI system is designed and its capability to achieve its intended purpose, that is its “performance”¹⁴². Whereas the original

¹³⁴ European Commission, *Communication from the Commission. Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)*, Brussels, 29.7.2025, C(2025) 5052 final (hereinafter “Guidelines”).

¹³⁵ AI Act, Article 96(1)(b). “They should serve as practical guidance to assist competent authorities under the AI Act in their enforcement activities, as well as providers and deployers of AI systems in ensuring compliance with their obligations under the AI Act”. *Guidelines*, cited, 1.

¹³⁶ As the Guidelines recall the sole Institution competent to adopt an authoritative interpretation of such provisions is the CJEU.

¹³⁷ High-risk AI systems represent the core of the AI Act. See COTINO HUESO, *Scope And Delimitation Of High-Risk Systems In The Artificial Intelligence Act*, cited, 301-326.

¹³⁸ High-risk AI systems are not outright forbidden, but they do present significant risks to health, safety, and fundamental rights. Those rights include: right to human dignity, respect for private and family life, protection of personal data, freedom of expression and information, freedom of assembly and of association, the right to non-discrimination, the right to education, consumer protection, workers’ rights, the rights of persons with disabilities, gender equality, intellectual property rights, the right to an effective remedy and to a fair trial, the right of defence and the presumption of innocence, and the right to good administration. In addition to those rights, the AI Act recognise the rights of children and the fundamental right to a high level of environmental protection. AI Act, Recital 48.

¹³⁹ Annex I provides a list of Union harmonisation legislation, including both legislations based on the New Legislative Framework (e.g. machinery, toy safety, medical devices) and other legislation (e.g. agricultural and forestry vehicles, rail systems, two- or three-wheel vehicles, quadricycles). For a discussion on this first category see COTINO HUESO, *Scope And Delimitation Of High-Risk Systems In The Artificial Intelligence Act*, cited, 304-308.

¹⁴⁰ MARCHETTI, *Artificial Intelligence And Public Authorities: Does The European Ai Act Protect Public Values?*, cited, 77.

¹⁴¹ AI Act, Article 3(12).

¹⁴² AI Act, Article 3(18).

draft established an automatism, classifying any AI system as high-risk if it was intended for the purposes set out in Annex III, the final version removes this automatic link through the introduction of Article 6(3). This provision acknowledges that not every AI system deployed in a high-risk context should automatically be classified as high-risk. Where an AI system is limited in scope, auxiliary in nature, or merely preparatory, subjecting it to the full suite of high-risk requirements or compliance obligations may not be proportionate¹⁴³.

Accordingly, the provision specifies four circumstances in which the AI system does not pose a “*significant risk*” and thus should not be considered high-risk. First, the high-risk “status” does not apply where the AI system is designed to perform a narrow procedural task¹⁴⁴. In such cases, the system undertakes a simple, mechanical operation without exercising discretion. Within the law enforcement context, for instance, an AI system that automatically redacts personal identifiers (such as phone numbers or addresses) from police reports prior to their transmission to another agency would fall into this category. The function is purely routine and procedural, with no bearing on evaluative judgments such as determinations of suspicion. Second, the AI system is employed to improve the result of a previously completed human activity¹⁴⁵. Here, the system enhances but does not alter the substance of human input. An illustrative case would be an AI system that, following the drafting of an incident report by a police officer, corrects grammar, adjusts formatting, or inserts keyword tagging to facilitate subsequent database retrieval. In this instance, the officer’s substantive narrative remains intact, with the AI system contributing only to clarity and usability. Third, Article 6(3) considers AI systems that detect decision-making patterns or deviations without substituting human judgment¹⁴⁶. Such systems provide analytical support, highlighting trends while refraining from prescriptive recommendations. For example, an AI system might analyse past bail decisions to identify statistical outliers for internal audit purposes or review historical stop-and-search records to detect disproportionate targeting of certain groups. Fourth, the carve-out applies where the AI system performs a preparatory task in relation to an assessment relevant to Annex III use cases¹⁴⁷. In such scenarios, the system organizes or processes data to support, but not replace, subsequent human evaluation. For instance, an AI application that sorts CCTV footage chronologically and identifies all instances of a specific object to expedite investigative review, or one that clusters intercepted communications by language or topic

¹⁴³ AI Act, Article 6(4) applies: “A provider who considers that an AI system referred to in Annex III is not high-risk shall document its assessment before that system is placed on the market or put into service. Such provider shall be subject to the registration obligation set out in Article 49(2). Upon request of national competent authorities, the provider shall provide the documentation of the assessment”.

¹⁴⁴ AI Act, Article 6(3)(a).

¹⁴⁵ AI Act, Article 6(3)(b).

¹⁴⁶ AI Act, Article 6(3)(c).

¹⁴⁷ AI Act, Article 6(3)(d).

before human analysis, would fall under this provision. These tools function as organisational and triage mechanisms, while the decisive evaluation – whether relating to charges, investigations, or surveillance measures – remains firmly in human hands.

While the rationale behind this provision appears clear, it is not – from my perspective – entirely convincing. First, although the provision is formally addressed to AI providers, its practical application depends primarily on how the deployer uses the system¹⁴⁸. Second, even where these systems serve a preparatory function, they may still influence decision-making, for instance, through information selection or categorisation. Third, while providers that apply the filter are required to register their decision in the EU database pursuant to Article 6(4) of the AI Act, thereby providing some overview of how the filter is applied in practice, Article 49(4) introduces exceptions in the area of law enforcement that limit the transparency of these registrations¹⁴⁹. In any case, the AI Act explicitly maintains that systems listed in Annex III are always to be considered high-risk where they involve the profiling of natural persons, irrespective of whether they otherwise meet the conditions for exemption under Article 6(3). Person-based predictive policing, which typically relies on profiling, is therefore excluded from the scope of Article 6(3).

Once an AI system is classified as high-risk, it must comply with the essential requirements, and both the provider and the deployer must fulfil the obligations set out in Chapter III of the AI Act. The fundamental principle underpinning the AI Act is that those engaged in the development or deployment of AI systems bear responsibility for managing risks across all phases of the AI lifecycle discussed in Chapter 1.

Providers: Articles 9-22

Article 9	Risk Management System
Article 10	Data and Data Governance
Article 11	Technical Documentation
Article 12	Record-Keeping
Article 13	Transparency and Provision of Information to Deployers
Article 14	Human Oversight
Article 15	Accuracy, Robustness and Cybersecurity
Article 16	Obligations of Providers of High-Risk AI Systems
Article 17	Quality Management System
Article 18	Documentation Keeping
Article 19	Automatically Generated Logs
Article 20	Corrective Actions and Duty of Information
Article 21	Cooperation with Competent Authorities

¹⁴⁸ In similar terms COTINO HUESO, *Scope And Delimitation Of High-Risk Systems In The Artificial Intelligence Act*, cited, 322.

¹⁴⁹ FRA, *Assessing High-Risk Artificial Intelligence. Fundamental Rights Risks*, Report, Vienna, 2025.

CHAPTER 3

Article 22	Authorised Representatives of Providers of High-Risk AI Systems

Deployers: Article 26

Paragraph 1	Technical and organisational measures to ensure the use the AI system in accordance with instructions for use
Paragraph 2	Human oversight with the necessary competence, training and authority, as well as the necessary support
Paragraph 4	Quality of input data
Paragraph 5	Monitoring of operations of high-risk AI systems and information to providers
Paragraph 6	Record-keeping
Paragraph 8	Registration of the high-risk AI system in the EU database
Paragraph 10	Specific rules for post-remote biometric identification
Paragraph 11	Duty to inform the natural person concerned
Paragraph 12	Cooperation with Competent Authorities

For this section of the AI Act, as anticipated, the European legislator opted for a structure in line with the New Legislative Framework, which includes the need for AI systems to comply with technical requirements, the obligation of providers to undergo a conformity assessment procedure¹⁵⁰ and carry out post-market monitoring¹⁵¹.

3.3. The by-design regulation approach

Regulation by design is an emerging and rapidly expanding field of research, especially in the context of digital technologies¹⁵², that explores how legal and regulatory requirements can be embedded directly into the design of technologies, systems, and processes. The concept of regulation by design in the European landscape is not entirely new; it is well established in the field of data protection, most prominently through the principle of *privacy by design*. Enshrined in instruments such as the GDPR¹⁵³, privacy by design requires that data protection principles (such as data minimisation) are integrated

¹⁵⁰ ORTIGOSA, *Conformity Assessment In The Design And Production Of Artificial Intelligence-Based Systems In The Context Of The “New Legislative Framework”*, cited.

¹⁵¹ I. SALAZAR, M.Á. LIÉBANAS, *Post-Market Monitoring On High-Risk AI Systems In The Artificial Intelligence Act. Description, Measures And Use Cases*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 785-797.

¹⁵² K. PRIFTI, J. MORLEY, C. NOVELLI, L. FLORIDI, *Regulation by Design: Features, Practices, Limitations, and Governance Implications*, in *Minds and Machines*, 34(13), 2024, 1-23. M. ALMADA, *Regulation by Design and the Governance of Technological Futures*, in *European Journal of Risk Regulation*, 14(4), 2023, 697-709.

¹⁵³ L. FLORIDI, *Soft ethics, the Governance of the digital and the General Data Protection Regulation*, in *Philosophical Transactions of the Royal Society*, 2018. Available at: <https://royalsocietypublishing.org/doi/epdf/10.1098/rsta.2018.0081>.

into the development of technologies and organisational practices from the outset, rather than being treated as an afterthought¹⁵⁴. In practice, rather than relying solely on external oversight or *ex post* enforcement, regulation by design seeks to shape compliance *ex ante*, by ensuring that technologies are developed, from the beginning, in ways that inherently reflect legal norms, ethical principles, and regulatory objectives¹⁵⁵. To this end, legal provisions are directed primarily at the designers of digital technologies, who bear the responsibility of ensuring that their systems meet the essential requirements laid down in the legislation.

The European Union has adopted this approach in respect of several of the essential requirements governing high-risk AI systems, as set out in Chapter III, Section 2 of the AI Act¹⁵⁶:

Article 12(1): “High-risk AI systems shall *technically allow* for the automatic recording of events (logs) over the lifetime of the system” (Record-Keeping)

Article 13(1): “High-risk AI systems shall be *designed* and *developed* in such a way as to ensure that their operation is sufficiently transparent to enable deployers to interpret a system’s output and use it appropriately” (Transparency)

Article 14(1): “High-risk AI systems shall be *designed* and *developed* in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which they are in use” (Human Oversight)

Article 15(1): “High-risk AI systems shall be *designed* and *developed* in such a way that they achieve an appropriate level of accuracy, robustness, and cybersecurity, and that they perform consistently in those respects throughout their lifecycle” (Accuracy, Robustness and Cybersecurity)

Similar considerations can be stressed for limited-risk AI systems:

¹⁵⁴ One of the cornerstone principles of data protection legislation is the concept of ‘privacy by design,’ enshrined in Article 25 of the GDPR. On this topic see the interesting work of A. CAVOUKIAN, *Privacy by Design – The 7 Foundational Principles – Implementation and Mapping of Fair Information Practices*, May 2010. Available at <https://privacy.ucsc.edu/resources/privacy-by-design---foundational-principles.pdf>.

¹⁵⁵ “Under this approach, regulators oblige the designers of digital systems to ensure that their system design incorporates specific legal requirements. Designers, in turn, comply with such a requirement by turning the legal requirements into technical requirements which embed the contents of the RbD [regulation by design] provision into the code of the digital system”. ALMADA, *Regulation by Design and the Governance of Technological Futures*, cited, 697-698. In similar terms, see E. BAYAMLIOĞLU, R. LEENES, *The ‘Rule of Law’ Implications of Data-Driven Decision-Making: A Techno Regulatory Perspective*, in *Law, Innovation and Technology*, 10(2), 2018, 295-298. Methods used in regulation by design are described in PRIFTI, MORLEY, NOVELLI, FLORIDI, *Regulation by Design: Features, Practices, Limitations, and Governance Implications*, cited, 7.

¹⁵⁶ K. PRIFTI, *Regulating Artificial Intelligence by Design: Exploring the practice, developing the theory, and devising governance solutions*, PhD Thesis, Erasmus University Rotterdam, 2024. R. MAHARI, A. PENTLAND, *Regulation by design: a new paradigm for regulating AI systems*, in *Digital Single Market and Artificial Intelligence*, 2024. “Compliance and regulation by design represent a risk-management paradigm that’s uniquely suited for AI” Mahari said. “Intelligent technology design can proactively prevent failures and risks”. Available at: <https://blog.irvingwb.com/blog/2024/07/ai-is-uniquely-suited-to-regulation-by-design.html>.

Article 50(1): “Providers shall ensure that AI systems intended to interact directly with natural persons are *designed* and *developed* in such a way that the natural persons concerned are informed that they are interacting with an AI system” (Transparency)

From the European perspective, regulation by design may help in pursuing different positive goals¹⁵⁷. In the context of AI, regulation by design implies that public authorities, such as law enforcement authorities, would, at least in principle, procure and deploy AI systems that comply with the requirements laid down in the AI Act, since compliance with essential requirements is embedded within the technical architecture of AI systems before they are placed on the market, put into service, or used, ensuring that safeguards for fundamental rights are built into the system from the outset, and thereby affording individuals a higher level of protection.

Despite these advantages, regulation by design is not without its critics. A primary concern is the risk that designers may effectively assume the role of “rule-makers”¹⁵⁸. In order to comply with legal obligations, designers inevitably need to interpret the law and decide on the technical means to address its demands, for instance by hardcoding or soft coding a particular rule into the system¹⁵⁹. In support of this process, technical requirements set out in general terms within legislation are frequently given practical effect through standards developed by standardisation bodies, certification schemes, or a combination of such mechanisms. The AI Act follows the same approach (see Section 2.4), requiring European standardisation organisations to develop the necessary technical specifications and thus delegating certain practical implementation responsibilities to those bodies¹⁶⁰. While such delegation may simplify compliance with legal obligations, it raises concerns about legitimacy by transferring regulatory authority – *i.e.* the power to determine the actual content of the regulatory requirements – from democratically accountable lawmakers to standardisation organisations, private certification bodies or directly to designers¹⁶¹.

Technical concerns may arise since translating normative concepts into technical rules is not always straightforward¹⁶². The formulation or content of a legal rule may represent an obstacle to its

¹⁵⁷ C. DJEFFAL, *Law by Design Obligations: The Future of Regulating Digital Technologies in Europe?*, in *Computer Law & Security Review: The International Journal of Technology Law and Practice*, 59, 2025, Article 106232, 1-18. Available at SSRN: <https://ssrn.com/abstract=4765471> or <http://dx.doi.org/10.2139/ssrn.4765471>. See also PRIFTI, MORLEY, NOVELLI, FLORIDI, *Regulation by Design: Features, Practices, Limitations, and Governance Implications*, cited, 4 where the authors distinguish between “high-level goals”, such as democracy, human rights, and the rule of law, and “low-level goals”, such as contestability, explainability, and security.

¹⁵⁸ In these terms, ALMADA, *Regulation by Design and the Governance of Technological Futures*, cited, 699.

¹⁵⁹ *Ibidem*.

¹⁶⁰ AI Act, Article 40.

¹⁶¹ In these terms see ALMADA, *Regulation by Design and the Governance of Technological Futures*, cited, 698. The author talks about regulation by design as “technological co-regulation”.

¹⁶² “The third challenge for RBD [regulation by design] is the methodological and epistemological problem of operationalising open-ended normative concepts (e.g., ethical principles) into workable solutions” PRIFTI, MORLEY, NOVELLI, FLORIDI, *Regulation by Design: Features, Practices, Limitations, and Governance Implications*, cited, 14.

expression in a programming language¹⁶³. Legal rule may contain principles – such as the principle of fairness – which are difficult to implement into AI systems, since its content is context-dependent and open to interpretation. Furthermore, legal rules often require the balancing of competing interests, whereas technical systems tend to hardcode a single solution, leaving little room for contextual judgment, a challenge that becomes particularly acute when the underlying value judgments shift over time. Consider, for instance, the obligation imposed on social media platforms to automatically remove hate speech¹⁶⁴. While automated filters can detect many unlawful posts, they also risk wrongful removals, for example, where content constitutes parody, satire, political critique, or artistic expression. Such moderation errors can significantly infringe on users’ fundamental rights, most notably freedom of expression. Whether a removal is justified depends not only on the context (it is a joke, or a legitimate form of protest) but also on cultural norms that define what is acceptable in that society. Designers cannot anticipate all such factors in advance, and even well-calibrated systems risk obsolescence as societal attitudes toward certain speech evolve.

Another issue relating to regulation by design is the cost of compliance. Smaller firms and start-ups, in particular, struggle to integrate regulatory requirements into their systems, especially in an ever-evolving sector such as that of AI¹⁶⁵. Finally, the absence of clear, measurable compliance metrics to evaluate the compliance with legal requirements is another issue¹⁶⁶. The accuracy requirement can be taken as an example. Article 15 states that high-risk AI systems must be designed and developed to achieve an “*appropriate level of accuracy*”, but it does not specify what this means in practice. Designers thus have ample leeway in selecting the techniques used to achieve compliance, provided that all essential requirements of the AI Act are met in full and simultaneously¹⁶⁷. Given the technical challenges of measuring appropriate levels of accuracy, the Commission, in cooperation with relevant

Computer scientists have developed various techniques to represent legal requirements in software. For an introduction to such techniques see T. BENCH-CAPON, *Thirty Years of Artificial Intelligence and Law: Editor’s Introduction*, in *Artificial Intelligence and Law*, 30, 2022, 475–479. The translation may be easy when rules are formulated in a conditional structure (“if A, then B”), similar to the conditional logic in programming language.

¹⁶³ For an interesting discussion on the changes that the law is living due to its translation in a digital language see M.R. FERRARESE, *Digitalizzazione e diritto. Alcune domande e questioni*, cited, 23-37. Garapon and Lassègue defined the digital language as “*mute and opaque*”. See A. GARAPON, J. LASSÈGUE, *La giustizia digitale*, il Mulino, 2021, 49.

¹⁶⁴ G. SARTOR, A. LOREGGIA, *The Impact of Algorithms for Online Content Filtering or Moderation. Upload Filters*, Study commissioned by the European Parliament’s Policy Department for Citizens’ Rights and Constitutional Affairs, 2020.

¹⁶⁵ PRIFTI, MORLEY, NOVELLI, FLORIDI, *Regulation by Design: Features, Practices, Limitations, and Governance Implications*, cited, 15-16.

¹⁶⁶ R. MAHARI, A. PENTLAND, *Regulation by design: a new paradigm for regulating AI systems*, cited, 437-438.

¹⁶⁷ Selecting the most accurate model does not always equate to choosing the most appropriate solution, as gains in predictive performance may come at the expense of explainability, or other essential requirements under the AI Act. It is therefore necessary to conduct a careful assessment of trade-offs – such as accuracy versus interpretability, complexity versus robustness, or efficiency versus fairness – in order to identify the option that best balances technical performance with legal and ethical obligations. For a discussion on trade-offs between accuracy and explainability see A. BELL ET AL., *It’s Just Not That Simple: An Empirical Study of the Accuracy–Explainability Trade-off in Machine Learning for Public Policy*, ACM Conference on Fairness, Accountability, and Transparency, Seoul, 2022.

stakeholders and organisations, is called to encourage the development of benchmarks and measurement methodologies¹⁶⁸.

3.4. *The co-regulatory approach: the technical standards*

As anticipated, the European legislature opted to govern AI through a synergy of regulatory instruments. Despite the horizontal and risk-based approaches form the core of the AI Act's regulatory architecture, they do not exhaust the range of approaches employed by the European legislator. While the horizontal approach provides a general framework applicable across all sectors, and the risk-based approach enables differentiation based on the potential impact of specific AI systems, these methods alone are not sufficient to capture the full complexity and contextual nuances of AI applications. To enable a more targeted and use-specific implementation of requirements, therefore, the AI Act embraced also a co-regulatory approach¹⁶⁹.

As discussed above, the AI Act establishes regulatory requirements for AI systems, organised around a tiered risk-based framework, some of which are broad and indeterminate in scope. To ensure their consistent and technically feasible implementation thus the AI Act relies on harmonised standards and common specifications. While both instruments are formally voluntary¹⁷⁰, they carry significant "legal-public effect"¹⁷¹: AI systems developed in accordance with these technical documents are presumed to comply with the essential requirements established for high-risk AI systems under the AI Act.

Harmonised standards¹⁷² represent a specific category of technical documents that, reflecting the state of the art of the development of science and technology, are aimed at supporting the implementation

¹⁶⁸ AI Act, Article 15, par. 2.

¹⁶⁹ The role of standardisation in the EU digital and industrial strategy is evident from 2015. European Commission, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Better Regulation for Better Results-EU Agenda*, 19.05.2015, COM(2015) 215 final. The European Commission presented its new standardisation strategy in 2022. The European Standardization Bodies are currently developing standards across multiple sectors, including for the Digital Operational Resilience Act (DORA). In the field of AI, standardisation has been identified in the 2021 *Coordinated Plan* as a key area of joint action between the European Commission and the Member States. See M. CANTERO GAMITO, C.T. MARSDEN, *Artificial intelligence co-regulation? The role of standards in the EU AI act*, in *International Journal of Law and Information Technology*, 2024, 32(1), Article eaae011.

¹⁷⁰ This means that economic operators do not face fines for failing to comply with them. Nevertheless, such standards may carry a *de facto* obligatory character. See V. ÁLVAREZ GARCÍA, *El proceso de privatización de la calidad y de la seguridad industrial y sus implicaciones desde el punto de vista de la competencia empresarial*, in *Revista de Administración Pública*, 159, 2002, 344.

¹⁷¹ V. ÁLVAREZ GARCÍA, *The implementation of harmonised standards and common specifications*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 418.

¹⁷² M. ELIANTONIO, A. VOLPATO, *The European System of Harmonised Standards. Legal Opinion for ECOS*, Maastricht Faculty of Law Working Paper, 11 March 2022. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4055292.

of EU harmonisation legislation¹⁷³. Therefore, the legislative act lays down the essential requirements that products must comply with, and harmonised standards lay down the technical specifications. They are drawn up by the European Standards Organizations (ESOs)¹⁷⁴ following a complex process¹⁷⁵. Harmonised standards are initiated through a formal standardisation request issued by the European Commission¹⁷⁶. This request is initially drafted by the Commission, then refined in consultation with the ESOs and relevant stakeholders and finally adopted by the Commission. Once the standardisation request is finalised, the ESOs begin drafting the standards. At this stage, National Standards Bodies (NSBs)¹⁷⁷ play a central role, operating through technical committees that bring together representatives of all principal parties with a national interest. These stakeholders designate their representatives (referred to as Experts) who contribute to the development of standards within dedicated Working Groups. More specifically, Working Groups are responsible for the development of the first draft of the standard. Stakeholder feedback is then collected and reviewed through an official enquiry. Based on this input, the NSBs may revise the initial draft. Once the draft is either updated or validated, the European Commission assess, also with the help of external consultants, whether the document drafted by the ESO is compliant with its initial request and with the requirements set out in secondary legislation. If the standard is deemed compliant, it is put to a vote. Upon successful adoption, it is published in the Official Journal of the European Union, thereby gaining official status.

¹⁷³ Harmonised standards are only those which are developed to support EU legislation. Not all standards, however, have this specific aim. For example, if the ESO develop standards on its own initiative, they might not be harmonised standards. For a description of EU standardisation see Regulation (EU) No 1025/2012.

¹⁷⁴ AI Act, Article 40. The standardisation bodies of the European Union are the European Committee for Standardisation (CEN), the European Committee for Electrotechnical Standardisation (CENELEC) and the European Telecommunications Standards Institute (ETSI). The AI Act mandated the CEN and CENELEC to draft the standards relating to the AI Act. In the European Union, only standards developed by CEN, CENELEC and ETSI are recognised as European Standards. The ESOs are independent of the EU institutions. However, the European Commission can request ESOs to produce standards to support EU legislation, such as in the case of harmonised standards. At the international level, the standardisation setting organisations are the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC) and the International Telecommunication Union (ITU). At the national level, one of the most important standardisation organisations is the National Institute of Standards and Technology (NIST). The literature on standardisation is huge. Among the many works, see J.L. CONTRERAS, *The Cambridge Handbook of Technical Standardization Law. Volume 2: Further Intersections of Public and Private Law*, Cambridge University Press, 2019. M. CANTERO, M.W. MICKLITZ (EDS.), *The Role of the EU in Transnational Legal Ordering: Standards, Contracts and Codes*, Edward Elgar Publishing, 2020.

¹⁷⁵ For a detailed description of the standardisation process see Commission Staff Working Document, *Vademecum on European Standardisation in support of Union Legislation and policies: PART I Role of the Commission's Standardisation requests to the European standardisation organisations*, Brussels, 27.10.2015 SWD(2015) 205 final; *PART II Preparation and adoption of the Commission's standardisation requests to the European standardisation organisations*; *PART III Guidelines for the execution of standardisation requests*.

¹⁷⁶ Regulation (EU) No 1025/2012, Article 2: “‘harmonised standard’ means a European standard adopted on the basis of a request made by the Commission for the application of Union harmonisation legislation”.

¹⁷⁷ National Standards Bodies represent the interests of all stakeholders in their respective countries, including government, industry, and civil society, by allowing national-level parties to become members of their committees.

The legislator (the Council and Parliament) can challenge harmonised standards through a formal objection¹⁷⁸.

In relation to the AI Act, the European Commission shall issue a Standardisation Request that covers all the requirements and obligations set out in Chapter III Section 2 and in Chapter V, Sections 2 and 3. As requested by Article 40(2), when preparing a standardisation request, the Commission shall consult the AI Board and relevant stakeholders, including the advisory forum¹⁷⁹. In 2022, the European Commission issued a Standardisation Request¹⁸⁰ to the ESOs. Currently, close to one thousand Experts drawn from all European Member States are engaged in developing the standards required for the implementation of the AI Act, under the auspices of CEN-CENELEC JTC 21 “Artificial Intelligence”. The architecture of this standards framework is considerable in its complexity, comprising numerous documents each focused on one or more specific requirements¹⁸¹.

The AI Act’s reliance on standards has positive and negative implications¹⁸². Indeed, reliance on standards to support AI Act implementation appears necessary, as they complement the general requirements set out in the Regulation. Moreover, as noted above, harmonised standards carry a presumption of conformity with the harmonised legislation. In the specific context of the AI Act, this means that where an AI provider adheres to the applicable harmonised standards covering the relevant requirements, their AI system is presumed to comply with the Regulation, a mechanism that simplifies legal compliance and helps reduce the associated costs¹⁸³. In the case of biometrics, for instance,

¹⁷⁸ For a list of European Commission decisions taken on the basis of formal objections raised against publication of the references of harmonised standards in the Official Journal of the European Union see https://single-market-economy.ec.europa.eu/single-market/goods/european-standards/harmonised-standards/formal-objections_en.

¹⁷⁹ The Commission shall specify that “standards have to be clear, consistent, including with the standards developed in the various sectors for products covered by the existing Union harmonisation legislation listed in Annex I, and aiming to ensure that high-risk AI systems or general-purpose AI models placed on the market or put into service in the Union meet the relevant requirements or obligations laid down in this Regulation. The Commission shall request the European standardisation organisations to provide evidence of their best efforts to fulfil the objectives referred to in the first and the second subparagraph of this paragraph in accordance with Article 24 of Regulation (EU) No 1025/2012”.

¹⁸⁰ European Commission, *Draft standardisation request to the European Standardisation Organisations in support of safe and trustworthy Artificial Intelligence*, 5 December 2022. Available at: <https://ec.europa.eu/docsroom/documents/52376>.

¹⁸¹ Current standards in phase of development are: prEN 18286 “Quality management system for the EU AI Act regulatory purposes” (WI JT021039); prEN 18229-1 “AI trustworthiness framework – Part 1: Logging, transparency and human oversight” (WI JT021008); prEN 18229-2 “AI trustworthiness framework – Part 2: Accuracy and robustness” (WI JT021047); prEN 18228 “AI Risk Management” (WI JT021024); prEN 18281 “Evaluation methods for accurate computer vision systems” (WI JT021025); prEN 18282 “Cybersecurity specifications for AI systems” (WI JT021029); prEN 18283 “Concepts, measures and requirements for managing bias in AI systems” (WI JT021036); prEN 18284 “Quality and governance of datasets in AI” (WI JT021037); prEN 18285 “AI Conformity assessment framework” (WI JT021038).

¹⁸² J. SOLER GARRIDO, S. DE NIGRIS, E. BASSANI, I. SÁNCHEZ, T. EVAS, A.-A. ANDRÉ, T. BOULANGÉ, *Harmonised Standards for the European AI Act*, in JRC policy brief, JRC 139430. A central study on harmonised standards in the AI Act is M. MCFADDEN ET AL., *Harmonising Artificial Intelligence: The role of standards in the EU AI Regulation*, Oxford Commission on AI&Good Governance, December 2021.

¹⁸³ AI Act, Article 40. See MCFADDEN ET AL., *Harmonising Artificial Intelligence: The role of standards in the EU AI Regulation*, cited, 7. Of course, economic operators might demonstrate compliance also with technical solutions other than those described in the harmonised standards. Compliance with them, simply determines a legal presumption of conformity with the legal requirements of the AI Act.

compliance with harmonised standards or common specifications allows providers to opt for a conformity assessment procedure based on internal control, rather than being required to involve a notified body.

Even though standards play a crucial role in translating abstract legal obligations into precise, sector-specific, and technically detailed requirements, they present legitimacy issues in the European landscape¹⁸⁴. First, the standards produced by ESOs have traditionally been regarded as private in nature¹⁸⁵. However, the practice of EU institutions requesting technical specifications to complement legislative acts within the New Approach has progressively conferred a public-law function upon technical standards. In 2016, the Court of Justice of the European Union gave formal recognition to this development, holding that such standards constitute part of Union law¹⁸⁶.

Although this represents a relevant change in perspective, participation in the development of technical standards is often limited to industrial stakeholders, raising issues of inclusivity and appropriate representation¹⁸⁷. As research shows, in fact, the standards represent the result of a “*collaborative effort of an engaged set of stakeholders*” and those who are involved are generally “*engineers within industry, whose organisations have a direct economic stake in the outcome*”¹⁸⁸. Non-industrial stakeholders generally do not participate in the standardisation process due to a lack of resources; when they do, their participation is often limited, either without voting rights or, when they are able to vote, their influence tends to be marginal. Standardisation processes are therefore predominantly driven by industrial considerations, raising ongoing concerns about their ability to adequately address non-industrial aspects, such as social impact, public interest and protection of fundamental rights¹⁸⁹.

¹⁸⁴ M. ELIANTONIO, C. CAUFFMAN, *The Legitimacy of Standardisation as a Regulatory Technique - A Cross disciplinary and Multi-level Analysis: an introduction*, in M. ELIANTONIO, C. CAUFFMAN (Eds.), *The Legitimacy of Standardisation as a Regulatory Technique - A Cross disciplinary and Multi-level Analysis*, Edward Elgar Publishing., 2020. M. EBERS, *Standardizing AI – The Case of the European Commission’s Proposal for an Artificial Intelligence Act*, in L. A. DIMATTEO, M. CANNARSA, C. PONCIBÒ (Eds.), *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics*, Cambridge University Press, 2022, 321-344.

¹⁸⁵ “*European standardisation bodies are private subjects, acting according to private procedures, their standards have traditionally been private*”. ÁLVAREZ GARCÍA, *The implementation of harmonised standards and common specifications*, cited, 423. Despite the evolution in the case law of the CJEU, European standards are still private in nature.

¹⁸⁶ CJEU, Judgement of 27 October 2016, *James Elliott Construction Limited v. Irish Asphalt Limited*, case C-613/14, EU:C:2016:821. For a discussion on the CJEU’s judgement see V. ÁLVAREZ GARCÍA, *La confirmación por parte de la jurisprudencia del Tribunal de Justicia de la Unión Europea de la capacidad normativa de los sujetos privados y sus lagunas jurídicas (el asunto “James Elliott Construction Limited contra Irish Asphalt Limited”)*, in *Revista General de Derecho Administrativo*, 2017, 46.

¹⁸⁷ The participation of other stakeholders is requested by the law. See Regulation 1025/2012, Article 5. Their participation, however, depends on the procedures adopted by each ESO. In practice, the limited resources of non-industrial stakeholders make it extremely difficult for them to take part in the standards-development process. To address this, Regulation 1025/2012 introduced special provisions for Annex III organisations, providing funding to support their involvement. Building on this, in 2022 the European Commission presented its Standardisation Strategy, calling for greater participation of SMEs, civil society, and users.

¹⁸⁸ MCFADDEN ET AL., *Harmonising Artificial Intelligence: The role of standards in the EU AI Regulation*, cited, 9-10.

¹⁸⁹ MANTELERO, *The Artificial Intelligence Act: The European Legislator’s Response To The Challenges Of Artificial Intelligence*, cited, 50-51. The author identifies multiple reasons why the use of standards is not adequate for a duly

An effort to strengthen stakeholder participation – and therefore improving account for non-industrial considerations – can be observed in Recital 121 of the AI Act, which, although not reflected in Article 40, calls for a “*balanced representation of interests involving all relevant stakeholders in the development of standards, in particular SMEs, consumer organisations and environmental and social stakeholder*”¹⁹⁰. A principal concern, however, remains the insufficient involvement of bodies tasked with safeguarding fundamental rights and public interests in the development of harmonised standards, a gap that represents one of the most pressing challenges inherent in relying on standardisation as a governance tool in the field of AI¹⁹¹. Therefore, questions persist as to whether technical standards are well-suited to address the ethical and fundamental rights considerations arising from the use of AI systems, particularly in areas where the interests at stake are significant¹⁹².

As anticipated, the nature and role of harmonised standards have evolved over time within the European framework. Today, harmonised standards are regarded as part of European law, having the CJEU clarified that harmonised standards are adopted to support the implementation of secondary legislation¹⁹³. As a result, access to them should be ensured¹⁹⁴, and they may be the subject of preliminary rulings¹⁹⁵. Although harmonised standards themselves cannot be directly challenged through an action for annulment under Article 263 TFEU – since ESOs are not considered EU bodies – the European Commission’s decision to reference these standards in the Official Journal can be

consideration of fundamental rights protection. First, there are not fundamental rights experts involved in the standardisation process. Second, the standardisation system is opaque. Third, standards are usable in relation to processes characterised by constant and repetitive dynamics, which is not the case in relation to AI, where the same application might have different impacts, depending on the context in which the system is used, in which scenario or depending on the individuals involved (such as minors or vulnerable persons). However, conclusions might be different depending on how standards are considered. If standards are considered as “*a general methodological framework for risk management in the case of fundamental rights*”, then fundamental rights considerations might be adequately considered. For an example of this approach, see A. MANTELERO, *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*, IT&LAW 36, Springer 2022. Doubts on the adequacy of standards in the protection of fundamental rights are highlighted also by G. OLIVATO, *The AI Act: regulatory and technical challenges*, cited, 298.

¹⁹⁰ AI Act, Recital 121.

¹⁹¹ MCFADDEN ET AL., *Harmonising Artificial Intelligence: The role of standards in the EU AI Regulation*, cited, 19.

¹⁹² M. CANTERO GAMITO, *The Role of ETSI in the EU’s Regulation and Governance of Artificial Intelligence*, 23 March 2024, *Innovation: The European Journal of Social Science Research*. Available at SSRN: <https://ssrn.com/abstract=4770324> or <http://dx.doi.org/10.2139/ssrn.4770324>.

¹⁹³ C. COLOMBO, M. ELIANTONIO, *Harmonized technical standards as part of EU law: Juridification with a number of unresolved legitimacy concerns?: Case C-613/14 James Elliot Construction Limited v. Irish Asphalt Limited*, EU:C:2016:821, in *Maastricht Journal of European and Comparative Law*, 2017, 24(2), 323-340. Available at: <https://doi.org/10.1177/1023263X17709753>. H. Schapel, *The New Approach to the New Approach: The Juridification of Harmonized Standards in EU Law*, in *Maastricht Journal of European and Comparative Law*, 2013, 20(4), 521-533. <https://doi.org/10.1177/1023263X1302000404>. For an interesting discussion of the legal effects of harmonised standards see A. VOLPATO, *The legal effects of harmonised standards in EU law: From hard to soft law, and back?*, in *Review of European Administrative Law*, 16(2), 2023, 5-35.

¹⁹⁴ Currently, harmonised standards have official publication restricted to their references (*i.e.*, their numerical codes and titles). Unlike harmonised standards, common specifications must be published in full in the Official Journal of the European Union (Article 297(2) TFEU).

¹⁹⁵ The reference goes to preliminary ruling for the interpretation of EU law. It is still not clear if they can be the object of preliminary rulings for validity.

contested. Regulation (EU) 1025/2012 has tightened the Commission's supervisory role, yet it does not define the intensity or depth of that control. However, given the evolving nature of harmonised standards, a more substantial and comprehensive oversight by the Commission appears increasingly necessary¹⁹⁶.

One of the main challenges of standardisation under the AI Act lies in the relationship between European and international standards. This relationship is far from straightforward, as European standards must also reflect and safeguard European values, trust and ethics aspects, elements that are not necessarily embedded in international standards. While duplicating existing standards would be counterproductive¹⁹⁷, limiting European standardisation activities to areas not covered internationally and relying primarily on international standards may, in practice, reduce the level of protection afforded to individuals¹⁹⁸.

Apart from the reliance on standards, Article 41 AI Act empower the European Commission to adopt common technical specifications¹⁹⁹ through implementing acts²⁰⁰ when harmonised standards are either unavailable or considered inadequate²⁰¹. Like harmonised standards, common specifications may be adopted to address the requirements and obligations set out in Section 2 of Chapter III, as well as Sections 2 and 3 of Chapter V. However, unlike harmonised standards, common specifications are drafted directly by the European Commission, following consultation with the advisory forum²⁰² and notification of the committee referred to in Article 22 of the 2012 Standardisation Regulation, composed of Member State representatives and chaired by the Commission. Unlike harmonised standards, therefore, common specifications are public technical documents. Common specifications share with harmonised standards the effect of creating a presumption of conformity for high-risk AI

¹⁹⁶ Regulation 1025/2012, Article 10 par. 5 and 6. A. VOLPATO, M. ELIANTONIO, *The Contradictory Approach of the CJEU to the Judicial review of Standards: A Love-Hate relationship?* in M. ELIANTONIO, C. CAUFFMAN (Eds.), *The Legitimacy of Standardisation as a Regulatory Technique: A Cross-disciplinary and Multi-level Analysis*, cited, 91-109.

¹⁹⁷ Moreover, the Vienna and Frankfurt Agreements concluded between the European standardisation bodies (CEN and CENELEC) and their international counterparts (ISO and IEC) grant priority to the standardisation work of the ISO/IEC system over that of CEN/CENELEC.

¹⁹⁸ ÁLVAREZ GARCÍA, *The Implementation Of Harmonised Standards And Common Specifications In The Field Of Artificial Intelligence (Articles 40 And 41 AIA)*, cited, 434.

¹⁹⁹ AI Act, Article 41. Recital 61. AI Act, Article 3 paragraph 28: “a document, other than a standard, containing technical solutions providing a means to comply with certain requirements and obligations established under this Regulation”. Another field where common specifications exist is medical devices. Regulations on Medical Devices (Regulation (EU) 2017/745) and on In-Vitro Diagnostic Devices (Regulation (EU) 2017/746).

²⁰⁰ TFEU, Article 291.

²⁰¹ AI Act, Recital 121: “The common specification should be an exceptional fall back solution to facilitate the provider's obligation to comply with the requirements of this Regulation, when the standardisation request has not been accepted by any of the European standardisation organisations, or when the relevant harmonised standards insufficiently address fundamental rights concerns, or when the harmonised standards do not comply with the request, or when there are delays in the adoption of an appropriate harmonised standard”. The priority of harmonised standards over common specification is clear also from Article 41, paragraph 4, AI Act.

²⁰² Which has a varietal composition.

systems and general-purpose AI models that comply with them. Providers who choose not to follow the common specifications must instead furnish a detailed justification, demonstrating that their technical solutions meet the relevant requirements to an equivalent or higher standard²⁰³.

3.5. Considerations on the overall AI Act's structure

The question underlying the preceding discussion is whether the architecture of the AI Act is adequate to ensure the effective protection of public values safeguarded by EU law. As previously noted, the AI Act is fundamentally conceived as a piece of product safety legislation. This means that, like other EU legislative acts governing products, it primarily regulates the conditions under which AI systems can be placed on the market, put into service, and used within the European territory. This product-law nature carries certain limitations, since it is typically designed to ensure technical compliance, safety, and market harmonization, rather than to serve as a comprehensive instrument for protecting fundamental rights or broader societal and public interests. As a result, while the AI Act seeks to safeguard health, safety, and fundamental rights, its effectiveness in doing so is inevitably influenced by the structural constraints of this legal model²⁰⁴.

Despite these limitations, the European legislator has made a deliberate effort to elevate the level of protection of individuals. The AI Act introduces a complex regulatory architecture that blends multiple approaches, seeking to strike a balance between multiple interests, namely market integration, innovation, and the preservation of public values protected by EU law.

As previously observed, central to this structure is the horizontal approach, considered the most suitable solution to ensure regulatory coherence and consistency across the single market and to avoid regulatory fragmentation. However, this approach is not without its limitations. As observed earlier, a single, overarching regulation of AI systems – one that does not differentiate between the techniques that qualify a system as an AI system under the AI Act, as discussed in Chapter 1, nor introduce specific obligations tailored to the purposes and contexts in which such systems are deployed – may prove problematic²⁰⁵. A horizontal framework, applied uniformly, risks either over-regulating low-risk applications or failing to capture harms that are specific to particular domains. The effectiveness of the

²⁰³ AI Act, Article 41, paragraph 5.

²⁰⁴ M. ALMADA, N. PETIT, *The EU AI act : a medley of product safety and fundamental rights?*, EUI, RSC, Working Paper, 2023/59. Available at: <https://hdl.handle.net/1814/75982>; M. VEALE, F. ZUIDERVEEN BORGESIU, *Demystifying the Draft EU Artificial Intelligence Act*, in *Computer Law Review International*, 22(4), 2021, 97-112, Available at SSRN: <https://ssrn.com/abstract=3896852>. M. GORNET, W. MAXWELL, *The European approach to regulating AI through technical standards*, in *Internet Policy Review*, 13(3), 2024. DOI: 10.14763/2024.3.1784. <https://policyreview.info/articles/analysis/regulating-ai-through-technical-standards>.

²⁰⁵ In similar terms, MARCHETTI, *Artificial Intelligence And Public Authorities: Does The European Ai Act Protect Public Values?*, cited, 73.

horizontal approach therefore depends significantly on the quality of sectoral guidance and complementary instruments, elements that the AI Act provides only in part. This is particularly true in relation to sensitive domains where the risks are heightened, and the ethical stakes are considerable. In these areas, a sector-specific regulatory approach may still be required to adequately address domain-specific challenges and ensure that fundamental rights and public trust are effectively safeguarded.

As far as the risk-based approach is concerned, the AI Act's stratification of obligations according to risk level represents a proportionate and pragmatically sound regulatory choice²⁰⁶. Nevertheless, the risk categorisation framework seems largely static and predetermined, which raises questions about its adaptability to technological evolution. Moreover, the boundaries between risk categories are not always clearly delineated, and the criteria for classification may give rise to legal uncertainty, particularly in cases where a system's intended purpose is ambiguous or where it migrates across use cases over time. Another critical aspect is the AI Act's approach to defining risk, reflecting choices that are normative as much as technical. Indeed, risk is not a neutral tool for rational regulation, but a politically charged concept shaped by market visions and power dynamics²⁰⁷. Moreover, the AI Act employs vague and ambiguous definitions of risk and does not provide clear methodologies for its identification or assessment. In particular, the AI Act regulates AI risks by using metrics, like the "*likelihood of [...] harm or adverse impact*", or "*severity of the harm that an AI system can cause*". While it introduces quantifiable metrics for risk assessment, this rationalised framework masks underlying normative decisions. Therefore, while such metrics may appear scientific and objective, they conceal value judgments that are inevitably embedded within them. For example, if an RBI system misidentifies individuals, what error rate should be considered "acceptable"? One per cent? Five per cent? Ten per cent? At what point does the error become "severe"? Who determines this threshold? Some authors have stressed that the AI Act uses *acceptability* as the standard. Therefore, if an AI system poses a risk that is judged "acceptable" then the AI system is considered "trustworthy"²⁰⁸. However, the AI Act never defines what "acceptable" means. Terms like "acceptable"²⁰⁹,

²⁰⁶ See, for example, M. EBERS, V.R.S. HOCH, F. ROSENKRANZ, H. RUSCHEMEIER, B. STEINRÖTTER, *The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS)*, in J, 4(4), 2021, 601. Available at: <https://www.mdpi.com/2571-8800/4/4/43>.

²⁰⁷ R. PAUL, *European artificial intelligence "trusted throughout the world": Risk-based regulation and the fashioning of a competitive common AI market*, in Regulation and Governance, 2024, 18(4), 1067. See L. BALLOT JONES, J. THORNTON, D. DE SILVA, *Limitations of risk-based artificial intelligence regulation: a structuration theory approach*, in *Discover Artificial Intelligence*, 2025, 4. Available at: <https://link.springer.com/article/10.1007/s44163-025-00233-9>.

²⁰⁸ J. LAUX, S. WACHTER, B. MITTELSTADT, *Trustworthy artificial intelligence and the European Union AI act: on the conflation of trustworthiness and acceptability of risk*, in Regulation and Governance, 18, 2024, 3–32.

²⁰⁹ AI Act, Article 9, paragraph 5: "*The risk management measures referred to in paragraph 2, point (d), shall be such that the relevant residual risk associated with each hazard, as well as the overall residual risk of the high-risk AI systems is judged to be acceptable*".

“appropriate”²¹⁰ or “commensurate”²¹¹ are vague and left to be interpreted by standardisation bodies and AI providers, effectively shifting substantive ethical decisions from lawmakers to private actors (*i.e.* standardisation bodies or AI providers). The problem is that, although the AI Act looks rational and scientific with its risk-based framework this approach presents issues²¹²: it reduces ethics to numbers and thresholds (“metrics”), without addressing deeper harms; it does not properly account for the context or the long-term effects of AI on society and it risks leaving fundamental ethical questions to private actors rather than democratic decision-making²¹³.

The AI Act’s requirement that certain properties – including transparency, robustness, accuracy, and human oversight – be embedded into AI systems from the outset rather than retrofitted after deployment is conceptually well-founded. It reflects an understanding that the architecture of an AI system fundamentally shapes its outputs and its potential for harm, and that meaningful accountability cannot be achieved through post-hoc intervention alone. Nonetheless, the by design approach places considerable technical and organisational demands on providers at an early stage in the development cycle. More fundamentally, a number of the properties it mandates, most notably fairness and non-discrimination, involve value judgments that resist reduction to engineering specifications or technical solutions. These properties are inherently contested and context-dependent, and their realisation typically requires normative balancing rather than technical optimisation. By attributing responsibility for safeguarding fundamental rights to AI developers, the by design framework risks shifting ethical and regulatory decision-making away from democratically accountable institutions and toward private actors whose primary mandate is commercial rather than public.

This dynamic is further compounded by the incentive structure created by harmonised standards. Where adherence to applicable standards carries a presumption of conformity, providers may be incentivised to apply those standards mechanically, treating compliance as an end in itself rather than as a means of giving genuine effect to the underlying legal and ethical requirements. The result risks a culture of superficial compliance, in which the by design obligations are approached as tick-box exercises rather than occasions for substantive and reflective engagement with the ethical and societal implications of the systems being developed.

²¹⁰ AI Act, Articles 9, 10, 12, 13, 14, 15.

²¹¹ AI Act, Article 14.

²¹² In these terms see also BALLOT JONES, THORNTON, DE SILVA, *Limitations of risk-based artificial intelligence regulation: a structuration theory approach*, cited.

²¹³ Nevertheless, the introduction of an obligation for AI providers to establish a risk management system, together with an obligation for AI deployers to conduct a fundamental rights impact assessment, may help to address these shortcomings. On the difference between these two requirements/obligations see E. CHAVELI DONET, *The Fundamental Rights Impact Assessment By Deployers Of Artificial Intelligence Systems In The Regulation*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 526-534.

CHAPTER 3

Taken together, the four pillars form a coherent regulatory architecture: the horizontal and risk-based approaches establish the framework's scope and structure; the by-design approach specifies the manner in which AI systems must be developed; and technical standards are tasked with operationalising all of the above. In this author's assessment, however, the structural weakness in this chain lies at its final link. Technical standards are asked to bear the full weight of translating broad, indeterminate, and value-laden legal requirements into concrete, measurable technical criteria, a function that goes well beyond the role for which standardisation processes were originally designed, and for which their existing governance structures may be insufficiently equipped. As discussed above, technical standards are neither democratically overseen nor value-neutral²¹⁴. In addressing normative questions (such as which metrics should be used or what error rates are acceptable) standards inevitably embody value judgments that shape, in ways that are rarely transparent, the rights and interests of those affected by AI systems²¹⁵. Yet because standards are developed through multi-stakeholder processes, reaching consensus on complex ethical questions is often difficult, if not impossible, with the result that such judgments tend to be resolved quietly, through technical choices, rather than openly deliberated. This is compounded by a more structural limitation: standards were originally conceived to ensure technical interoperability and uniformity, not to safeguard fundamental rights or give effect to broader societal values²¹⁶. Asking them to perform both functions simultaneously risks overburdening an instrument that was not designed with constitutional or democratic accountability in mind.

²¹⁴ GORNET, MAXWELL, *The European approach to regulating AI through technical standards*, cited, 13.

²¹⁵ C. ORWAT, J. BAREIS, A. FOLBERTH, J. JAHNEL, C. WADEPHUL, *Normative challenges of risk regulation of artificial intelligence and automated decision-making*, 2022. Available at: <https://link.springer.com/article/10.1007/s11569-024-00454-9>.

²¹⁶ Such as addressing risks of discrimination, invasion of privacy, chilling effects on fundamental rights. See H. POUGET, R. ZUHDI, *AI and Product Safety Standards Under the EU AI Act*, 5 March 2024. Available at: <https://carnegieendowment.org/research/2024/03/ai-and-product-safety-standards-under-the-eu-ai-act?lang=en>.

Part II: Law enforcement in the AI Act

1. The AI Act's provisions in the law enforcement context

The focus of the present Thesis is to analyse how the AI Act regulates the use of AI in the context of law enforcement, with particular attention to whether the framework affords adequate protection to individuals in such a sensitive domain. In Chapter 1, I examined the concept of law enforcement as set out in the LED, noting that the AI Act, in Articles 3(45) and 3(46), essentially mirrors that definition, thereby ensuring conceptual consistency between the two legal instruments.

The regulatory framework established by the AI Act in this area is notably intricate. It distinguishes AI systems placed on the market, put into service, or used for law enforcement purposes according to several levels of risk: some applications are outright prohibited (Article 5); others are classified as high-risk and therefore subject to stringent compliance requirements (Article 6; Annex III, point 6); a further group is deemed to present only limited risk, and the remainder of AI systems fall within the minimal risk category. This categorisation reflects the AI Act's risk-based approach, yet in the law enforcement context it is complicated by the presence of multiple layers of derogations: not only does it set out exceptions to general prohibitions or obligations but also introduces "exceptions to the exceptions" creating a nested structure of conditional permissions.

From a rights-protection perspective, the cumulative effect of these layers is problematic. First, the introduction of multiple derogations increases legal uncertainty, making it difficult for individuals to know the exact scope of their protections. Second, the decision-making power shifts significantly towards law enforcement authorities and national legislators, whose proportionality assessments may vary widely. Finally, the "exception to the exception" structure risks normalising practices that the AI Act has initially framed as extraordinary, potentially undermining the very rights-based rationale of the specific provisions.

2. Predictive Policing in the AI Act

In Chapter 2, I discussed predictive policing, emphasising that contemporary predictive policing software does not rely on a single, uniform technological foundation. Instead, such software employs a variety of computational techniques to generate predictions. Some systems are grounded in machine learning methods, which identify patterns and correlations in large datasets through statistical inference and adaptive algorithms. Others are built on more classical approaches, such as rule-based models or

statistical regression techniques, which follow predefined parameters and deterministic rules rather than adaptive learning. As discussed in Chapter 1, the definition of an AI system within the Regulation is deliberately broad, encompassing a wide range of computational approaches, provided that the conditions set out in Article 3(1) are met. Law enforcement authorities using predictive policing software must, therefore, first determine whether the software qualifies as an AI system under the AI Act's definition, and then determine the risk category into which the system falls.

Provisions on predictive policing underwent numerous changes during the legislative process. In its original version²¹⁷, the AI Act classified the broad category of predictive policing as high-risk under Annex III, point 6, identifying four different cases. Among these, letters a), e), and f) were explicitly related to person-based predictive policing, while letter g) could also refer to place-based predictive policing²¹⁸. The latter reference was removed by the Council upon the adoption of its General Approach in December 2022²¹⁹, lowering the regulatory scrutiny applied to place-based predictive policing systems. This was deeply concerning, as such systems – though less directly intrusive than person-based predictive policing – can still pose serious risks to fundamental rights, as discussed in Chapter 2. Apart from this reclassification, it remained unclear how points a), e), and f) would differ in scope or application²²⁰. Nevertheless, the Council's General Approach retained their original formulation, with the exception of a significant addition: it expressly clarified that these tools are to be regarded as high-risk not only when deployed directly by law enforcement authorities, but also when employed by other entities acting on their behalf.

The European Parliament introduced several amendments to this classification²²¹, notably designating person-based predictive policing as a prohibited practice under Article 5, with the exception of AI systems referred to in Annex III, point 6, letter f). While the rationale for retaining this exception remains unclear, the decision to classify person-based predictive policing as prohibited aligns with the

²¹⁷ European Commission, *Proposal for a Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts*, Brussels, 21 April 2021, COM(2021) 206 final, 2021/0106(COD). Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206>. Hereinafter “Proposal”

²¹⁸ In similar terms, F. MIRÓ LLINARES, M. SANTISTEBAN GALARZA, *The regulation of predictive policing systems in the Artificial Intelligence Act*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 340-341.

²¹⁹ Council of the European Union, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - General approach*, 25 November 2022, Interinstitutional File: 2021/0106(COD). Available at: <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf>. Hereinafter “Council's General Approach”.

²²⁰ MIRÓ LLINARES, SANTISTEBAN GALARZA, *The regulation of predictive policing systems in the Artificial Intelligence Act*, cited, 340-341.

²²¹ European Parliament, *Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, 14 June 2023, COM(2021)0206 – C9-0146/2021 – 2021/0106(COD). Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html. Hereinafter “Parliament's Amendments”.

CHAPTER 3

European Parliament’s Resolution of 6 October 2021²²² and the joint position of the EDPB and EDPS in Opinion 5/2021²²³, both of which highlighted concerns about the use of systems that enable the profiling of natural persons, especially in high-stake contexts. With regard to Annex III, point 6, letter g), the European Parliament retained the original formulation. The final version of the AI Act departs from earlier drafts, integrating several innovations introduced by both the Council and the European Parliament²²⁴.

EUROPEAN COMMISSION’S PROPOSAL	COUNCIL’S GENERAL APPROACH	EUROPEAN PARLIAMENT’S AMENDMENTS	FINAL VERSION
Article 5 (Prohibited AI practices): No reference to predictive policing	Article 5 (Prohibited AI practices): No reference to predictive policing	Article 5 (Prohibited AI practices) lett. (d a): “the placing on the market, putting into service or use of an AI system for making risk assessments of natural persons or groups thereof in order to assess the risk of a natural person for offending or reoffending or for predicting the occurrence or reoccurrence of an actual or potential criminal or administrative offence based on profiling of a natural person or on assessing personality traits and characteristics, including the person’s location, or past criminal behaviour of natural persons or groups of natural persons”	Article 5 (Prohibited AI practices) lett. (d): “the placing on the market, the putting into service for this specific purpose, or the use of an AI system for making risk assessments of natural persons in order to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics; this prohibition shall not apply to AI systems used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity”
Annex III, par. 6 lett. (a) “AI systems intended to be used by law enforcement authorities for making individual risk assessments of natural persons in order to assess the risk of a natural person for offending or reoffending or the risk for potential victims of criminal offences”	Annex III, par. 6 lett. (a) “AI systems intended to be used by law enforcement authorities or on their behalf to assess the risk of a natural person for offending or reoffending or the risk for a natural person to become a potential victim of criminal offences”	Annex III, par. 6 lett. (a): Deleted	Annex III, par. 6, lett. (a) “AI systems intended to be used by or on behalf of law enforcement authorities, or by Union institutions, bodies, offices or agencies in support of law enforcement authorities or on their behalf to assess the risk of a natural person

²²² European Parliament, *Resolution of 6 October 2021 on Artificial Intelligence in criminal law and its use by law enforcement authorities in criminal matters*, 2020/2016(INI).

²²³ EDPB-EDPS, *Joint Opinion 5/2021*, cited.

²²⁴ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Available at: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>. Hereinafter “AI Act” or “Regulation”.

			becoming the victim of criminal offences” Annex III, par. 6, lett. d) “AI systems intended to be used by law enforcement authorities or on their behalf or by Union institutions, bodies, offices or agencies in support of law enforcement authorities for assessing the risk of a natural person offending or re-offending not solely on the basis of the profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680, or to assess personality traits and characteristics or past criminal behaviour of natural persons or groups”
Annex III, par. 6 lett. (e) “AI systems intended to be used by law enforcement authorities for predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 or assessing personality traits and characteristics or past criminal behaviour of natural persons or groups”	Annex III, par. 6, lett. (e) “AI systems intended to be used by law enforcement authorities or on their behalf to predict the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 or to assess personality traits and characteristics or past criminal behaviour of natural persons or groups”	Annex III, par. 6, lett. (e): Deleted	Annex III, par. 6, lett. (e) AI systems intended to be used by or on behalf of law enforcement authorities or by Union institutions, bodies, offices or agencies in support of law enforcement authorities for the profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of the detection, investigation or prosecution of criminal offences.
Annex III, par. 6 lett. (f) AI systems intended to be used by law enforcement authorities for profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of detection, investigation or prosecution of criminal offences”	Annex III, par. 6, lett. (f): “AI systems intended to be used by law enforcement authorities or on their behalf to profile natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of detection, investigation or prosecution of criminal offences”.	Annex III, par. 6, lett. (f): “AI systems intended to be used by or on behalf of law enforcement authorities or by Union agencies, offices or bodies in support of law enforcement authorities for profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of detection, investigation or prosecution of criminal offences or, in the case of Union agencies, offices or bodies, as referred to in Article 3(5) of Regulation (EU) 2018/1725”	
Annex III, par. 6 lett. (g): “AI systems intended to be used for crime analytics regarding natural persons,	Annex III, par. 6, lett. (g) Deleted	Annex III, par. 6, lett. (g) : “AI systems intended to be used by or on behalf of law enforcement authorities or by Union agencies, offices	

allowing law enforcement authorities to search complex related and unrelated large data sets available in different data sources or in different data formats in order to identify unknown patterns or discover hidden relationships in the data”

or bodies in support of law enforcement authorities for crime analytics regarding natural persons, allowing law enforcement authorities to search complex related and unrelated large data sets available in different data sources or in different data formats in order to identify unknown patterns or discover hidden relationships in the data”

2.1.Prohibited predictive policing

Article 5(1)(d) prohibits the placing on the market, putting into service, and use of AI systems designed to assess or predict the likelihood that a natural person will commit a criminal offence, when such assessment relies solely on profiling or on the evaluation of personality traits and characteristics. This provision addresses what I referred to in Chapter 2 as person-based predictive policing, that is, the use of data analytics to estimate the risk of an individual engaging in criminal activity. Place-based predictive policing is outside the scope of the prohibition, since it does not generally involve the assessment of a specific individual. However, the *Guidelines on Prohibited AI practices* adopted by the European Commission also consider the case in which an AI system carries out place-based predictive policing, taking the risk score of the location into account when profiling a person. In this case, the system should be considered person-based, and, in principle, be covered by the prohibition if all the conditions are met²²⁵.

The prohibition under Article 5(1)(d) applies to both providers and deployers. As discussed in Chapter 2, law enforcement authorities are the primary deployers of predictive policing systems²²⁶, typically acquiring these systems from private companies. In some instances, however, law enforcement authorities may develop the system in-house, in which case they are considered providers. Furthermore, as discussed in Chapter 1, the notion of law enforcement authorities extends beyond public bodies expressly mandated for law enforcement purposes (such as the police) to include any entity or body entrusted by law with the exercise of public authority and powers for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including safeguarding against and preventing threats to public security. Accordingly, the prohibition

²²⁵ *Guidelines*, cited, paragraph 213.

²²⁶ *Guidelines*, cited, paragraph 71.

also covers private actors, such as banking institutions, when they act under a legal obligation to detect money laundering offences pursuant to anti-money laundering legislation. In other words, private entities or bodies must be entrusted by law with specific law enforcement tasks, act on behalf of law enforcement authorities, or be subject to specific legal obligations in order to be covered by the prohibition.

For an AI system to fall within the scope of the prohibition, several cumulative conditions must be met. First, the system must be used to assess or predict the risk of a natural person committing a criminal offence. Second, the system must be applied in relation to natural persons. Its use in connection with undertakings, such as companies or other legal entities, therefore, does not fall within the prohibition. As clarified by Recital 42, the prohibition does not extend to AI systems used for risk analytics aimed at assessing the likelihood of financial fraud by undertakings, for instance through the analysis of suspicious transactions. Third, the risk assessment or prediction must rely *solely* on profiling of the natural person, or on an evaluation of that person's personality traits and characteristics. As far as the first condition is concerned, by expressly linking the provision's application to "*criminal offences*", the European legislator seems to narrow the scope of Article 5(1)(d) to acts classified as criminal²²⁷.

In this author's assessment this can be concerning for two main reasons. First, as discussed in Chapter 1, the classification of an act as "criminal" varies across Member States, since what is criminal in one jurisdiction might be an administrative or civil infringement in another. However, in all cases where offences are not directly qualified as criminal and regulated by Union law, the national qualification of offences shall be interpreted in line with the CJEU's jurisprudence and subject to the scrutiny of the Court²²⁸. Second, the exclusion of "public security" from the explicit scope of the provision may be quite problematic. As discussed in Chapter 1, public security is a broad concept under EU law, covering threats to the functioning of society, public order, or safety that may not necessarily amount to criminal offences (e.g., certain large-scale protests, cyber incidents, or mass migration pressures). If the provision's scope is limited to "criminal offences", preventive actions in the interest of public security may fall outside its reach unless the conduct in question is also criminalized. Accordingly, this formulation risks undermining the provision's practical effectiveness, since AI systems deployed to safeguard against or prevent threats to public security – contexts that are non-criminal yet equally high-impact – fall outside the scope of the prohibition entirely.

²²⁷ The final version moved away from the Parliament's amendments, which explicitly referred also to administrative offences. Cfr. Parliament's Amendments, Amendment 224.

²²⁸ *Guidelines*, cited, paragraph 75.

As far as the third condition is concerned, determining whether risk assessment or prediction relies *solely* on profiling of the natural person, or on an evaluation of that person's personality traits and characteristics is not straightforward and must be assessed on a case-by-case basis. By way of illustration, the European Commission's Guidelines refer to a hypothetical law enforcement authority deploying an AI system to predict criminal behaviour in relation to offences such as terrorism, relying exclusively on personal attributes such as age, nationality, address, vehicle type, and marital status.

With regard to the adverb "*solely*", in relation to Article 11 of the LED and Article 22 of the GDPR it has been observed that a decision is not based solely on automated processing when such processing remains a "*decisional support tool*", provided that the human decision-maker genuinely considers the merits of the result, rather than being blindly or automatically guided by the system's output²²⁹. Indeed, the exclusion of the prohibition when the AI-based system is used for decision support is evident from Article 5(1)(d) itself. This provision excludes from its scope of application AI systems used to support human assessment of involvement in criminal activity based on objective and verifiable facts directly related to criminal activity. The second subparagraph of Article 5(1)(d) therefore clarifies the first subparagraph of the provision, explicitly exemplifying when the decision cannot consider as being "*solely*" based on profiling²³⁰.

A different interpretation of the adverb "*solely*" risks unduly narrowing the scope of the prohibition's application. Where any form of human assessment is present, particularly where additional factors are considered, the resulting risk assessment or prediction would no longer qualify as being based *solely* on profiling or on the evaluation of personal characteristics and therefore falls outside the scope of the prohibition. This construction creates a potential circumvention mechanism: even a perfunctory or ineffective human review could be invoked to evade the prohibition. Cognisant of this risk, the European Commission's Guidelines clarified that "*any such other elements will have to be real, substantial and meaningful [...] to be able to justify the conclusion that the prohibition does not apply*"²³¹.

²²⁹ L.A. BYGRAVE, *Article 22 automated individual decision-making, including profiling*. In: *The EU General Data Protection Regulation (GDPR)*, Oxford University Press, 2020, 533; Article 29 Data Protection Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, 2017, 20–21; M. BRKAN, *Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond*, in *International Journal of Law and Information Technology*, 27(2), 2019, 101–102.

²³⁰ Similarly, the Guidelines, stress that "*the prohibition does not apply if the AI system is used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to that activity*". *Guidelines*, cited, 65.

²³¹ *Guidelines*, cited, 68–69. "*As to the content of the exclusion, one of its central elements is that the AI system is used to support human assessment, rather than involving the AI system itself making the risk assessment as occurs in the situations covered by the prohibition. However, for the exclusion to apply, that human assessment must, in addition, already be based on objective and verifiable facts directly linked to a criminal activity*". *Guidelines*, cited, 70.

Existing predictive policing systems generally incorporate some form of human assessment, thereby complicating the threshold determination of what constitutes *solely* automated decision-making. It is evident that the implications of such interpretive uncertainty are far from negligible. If an AI-based predictive policing system was not deemed to rely solely on profiling or the evaluation of personal traits and characteristics, it would no longer fall within the scope of Article 5. Instead, it would be reclassified as a high-risk AI system under Article 6 of the AI Act.

The AI Act's definition of "profiling" in Article 3(52) mirrors that of Article 4(4) GDPR, as also reflected in Article 3(4) LED²³², meaning the Article 29 Working Party's guidance on GDPR profiling is directly applicable to its interpretation under the AI Act²³³. For the purposes of Article 4(4) – and therefore Article 3(52) of the AI Act – profiling requires three conditions to be met²³⁴: first, it must be an automated form of processing; second, data processed must be personal data; third, the purpose of profiling must be to evaluate personal aspects relating to a natural person. Therefore, the concept of profiling is connected to the categorisation of individuals and to the application of a "profile" to evaluate personal aspects relating to a natural person²³⁵. This interpretation is supported by the 2010 Committee of Ministers of the Council of Europe's Recommendation on the protection of individuals with regard to automatic processing of personal data²³⁶. In this document, profiling is described as "*an automatic data processing technique that consists of applying a 'profile' to an individual, particularly in order to take decisions concerning her or him or for analysing or predicting her or his personal preferences, behaviours and attitudes*"²³⁷. Consequently, attempting to profile a natural person without making an evaluation falls outside the scope of the definition.

Profiling shall not be confused with automated decision-making under Article 22 of the GDPR or Article 11 of the LED. While these two concepts are related, they must be used differently since automated decision-making may, but does not necessarily, imply profiling²³⁸. In any case, even if the

²³² AI Act, article 3(52). GDPR, Article 4(4): "*'profiling' means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements*".

²³³ Similar considerations were made in relation to article 3(4) of the LED. Article 29 Data Protection Working Party, *Guidelines on automated individual decisions and profiling for the purposes of Regulation 2016/679*, 6 February 2018.

²³⁴ MIRÓ LLINARES, SANTISTEBAN GALARZA, *The regulation of predictive policing systems in the Artificial Intelligence Act*, cited, 344.

²³⁵ The Article 29 Data Protection Working Party has opined that a '*simple classification of individuals based on known characteristics such as their age, sex, and height does not necessarily lead to profiling*', unless this classification progresses to making predictions or drawing conclusions about an individual. Article 29 Data Protection Working Party, *Guidelines on automated individual decisions and profiling for the purposes of Regulation 2016/679*, cited, 7.

²³⁶ Committee of Ministers of the Council of Europe, Recommendation CM/Rec(2010)13, *The protection of individuals with regard to automatic processing of personal data in the context of profiling*, 23 November 2010.

²³⁷ *Ibidem*, 9.

²³⁸ See K. WIEDEMANN, *Profiling and (automated) decision-making under the GDPR: A two-step approach*, in *Computer Law & Security Review*, 45, 2022, Article 105670, 1-14.

processing does not amount to profiling, Article 5(1)(d) prohibits person-based predictive policing when risk assessments are based on the personality traits and characteristics of individuals, which represent a broad category of features relating to a specific individual²³⁹. Since person-based predictive policing systems generally operate through profiling, which constitutes a “predictive method” for identifying “unknown” individuals who may be of interest to law enforcement authorities²⁴⁰, such tools may fall within the scope of the prohibition, provided the remaining definitional requirements are satisfied.

The restriction of person-based predictive policing within the European territory is grounded in several compelling considerations. As demonstrated in Chapter 2, predictive policing – especially in its person-based form – poses profound risks to the protection of fundamental rights and freedoms. Recital 42 makes this explicit from the outset: in accordance with the presumption of innocence, individuals in the European Union must be assessed solely on the basis of their actual behaviour, rather than on AI-generated predictions. The threat to rights is further exacerbated by the preventive function of such tools. In these circumstances, interventions may be triggered without any reasonable suspicion supported by objective and verifiable facts that a person is engaged in criminal conducts. This concern is confirmed by the wording of Recital 42 itself, which refers to the “*likelihood of their offending*” and the occurrence of “*an actual or potential criminal offence*”. Furthermore, as previously noted, profiling relies on the automated processing of personal data to generate an “individual profile”. In practice, however, such profiles often constitute generalisations drawn from group characteristics rather than an accurate reflection of individual ones²⁴¹, thereby risking the treatment of a person as a mere representative of a group rather than as an autonomous individual²⁴². Ideally, predictive methods would centre on actual behaviour. In reality, however, the focus frequently extends beyond behaviour to visible attributes such as age, gender, or ethnicity²⁴³.

²³⁹ Examples of such characteristics and traits are described in AI Act, Recital 42.

²⁴⁰ FRA, *Guidance on preventing unlawful profiling now and in the future*, 2019, 16. Available at: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-preventing-unlawful-profiling-guide_es.pdf. The other purpose of profiling is “to identify known individuals based on intelligence concerning a specific individual”. *Ibidem*.

²⁴¹ “Profiling establishes general correlations that may not be true for each individual. Any given individual may be the ‘exception to the rule’”. FRA, *Guidance on preventing unlawful profiling now and in the future*, cited, 16.

²⁴² MIRÓ LLINARES, SANTISTEBAN GALARZA, *The regulation of predictive policing systems in the Artificial Intelligence Act*, cited, 345.

²⁴³ While profiling has long been employed as a tool by law enforcement authorities, there are circumstances in which individualised decisions remain indispensable. *Ibidem*. In this regard, Recital 42 is explicit: “*In line with the presumption of innocence, natural persons in the Union should always be judged on their actual behaviour. Natural persons should never be judged on AI-predicted behaviour based solely on their profiling, personality traits or characteristics [...] without a reasonable suspicion of that person being involved in a criminal activity based on objective verifiable facts and without human assessment thereof*”.

In conclusion, the AI Act aligns with data protection legislation by establishing constraints on the processing of personal data that qualifies as profiling²⁴⁴.

The relation between prohibition under Article 5(1)(c) and Article 5(1)(d) is also relevant. The former prohibits a practice known as “social scoring”²⁴⁵. In order for the AI system to fall within the prohibition three conditions must be simultaneously fulfilled. First, the practice must constitute the “placing on the market”, the “putting into service” or the “use” of an AI system. Second, the AI system must be intended or used for the evaluation or classification of natural persons or groups of persons over a certain period of time based on: (a) their social behaviour; or (b) known, inferred or predicted personal or personality characteristics. Third, the social score created must lead or be capable of leading to the detrimental or unfavourable treatment of persons or groups: (a) in social contexts that are unrelated to those in which the data was originally generated or collected; and/or (b) when the treatment is unjustified or disproportionate to their social behaviour or its gravity²⁴⁶.

The “scoring” of natural persons is not categorically prohibited under the AI Act. It may be permissible where it pursues a clearly defined purpose and is conducted in compliance with Union and national law²⁴⁷. One such recognised purpose is the prevention, detection, investigation, prosecution, or punishment of criminal offences. In these circumstances, scoring is admissible only insofar it does not fall within the prohibition under Article 5 of the AI Act and as any resulting detrimental treatment of individuals is both justified and proportionate, in line with the substantive and procedural safeguards established by Union and national criminal and police law.

²⁴⁴ Decisions based on profiling may indeed have discriminatory effects, as stated by the CJEU, Judgement of 7 December 2023, *OQ v Land Hessen*, Case C-634/21. ECLI:EU:C:2023:957. Profiling may be unlawful if it leads to discriminatory outcomes, leading to an unjustified differential treatment of individuals on the basis of protected characteristics, or if personal data is processed unlawfully. To prevent discrimination, data quality is crucial. For the processing of personal data to be lawful, a legal basis is required, the purpose must be valid, lawful and appropriate (legitimate aim), the collection and processing must be necessary and proportionate (so called “reasonable justification”). FRA, *Handbook on European non-discrimination law – 2018 edition*, 21 March 2018. Right to non-discrimination and privacy and data protection are the rights considered to be mainly affected by unlawful profiling by the FRA. Unlawful profiling may also have negative impacts on the effectiveness of profiling. It may undermine trust in the police and deteriorate the relationship between the police and communities. See FRA, *Guidance on preventing unlawful profiling now and in the future*, cited, 38. Profiling is deemed to have no objective and reasonable justification “if it does not pursue a legitimate aim or if there is not a reasonable relationship of proportionality between the means employed and the aim sought to be realised”. European Commission against Racism and Intolerance (ECRI), *General Policy Recommendation N°11 on combating racism and racial discrimination in policing*, 4 October 2007, paragraph 28.

²⁴⁵ For a discussion on Article 5(1)(c) see M.Á. PRESNO LINERA, *The Prohibition Of Artificial Intelligence Systems That Evaluate And Classify People Based On Data That Are Unrelated To The Context In Which They Were Generated And That Lead To Discrimination*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 241-255.

²⁴⁶ *Guidelines*, cited, paragraph 149.

²⁴⁷ AI Act, Recital 31. The prohibition “should not affect lawful evaluation practices of natural persons that are carried out for a specific purpose in accordance with Union and national law”. The Guidelines stress that “In other words, AI systems which evaluate or classify individuals for the purposes of generating a social score in a lawful manner and for a specific purpose in the related context as that in which the personal data used for the score were collected are not prohibited, provided that any detrimental or unfavourable treatment from using the score is justified and proportionate to the gravity of the social behaviour”. *Guidelines*, cited, paragraph 176.

2.2. Other categories of predictive policing

As discussed above, for an AI system to fall within the prohibition set out in Article 5, three conditions must be met cumulatively. Where these conditions are not fulfilled, and the risk assessment is not based exclusively on profiling, Article 6 of the AI Act applies, and AI systems are classified as high-risk²⁴⁸. An example is when the risk assessment is not carried out solely by automated means and the involvement of human intervention is therefore expected to mitigate potential impacts on fundamental rights and freedoms²⁴⁹. The European Commission's Guidelines exemplify multiple cases in which the above conditions are not met, and AI systems thus fall outside the prohibition²⁵⁰.

As anticipated, determining whether the risk assessment or prediction relies *solely* on profiling of the natural person, or on an evaluation of that person's personality traits and characteristics is far from straightforward. Accordingly, taking the predictive policing systems described in Chapter 2 as an example, the mere fact that Operation LASER aims, *inter alia*, to identify individuals at risk of committing violent crimes is not, in itself, sufficient to qualify it as a prohibited practice. Rather, such a system may fall within the high-risk category where the predictive output is subsequently subject to a human assessment that is "*real, substantial and meaningful*".

Determining in advance whether an AI system falls into the high-risk or prohibited category is thus far from straightforward. Indeed, this classification depends on whether human intervention is involved in the decision-making process, and, if so, on the manner in which that human assessment is performed. Crucially, both of these factors cannot be fully defined at the system design stage, as they ultimately depend on how the deployer uses the system²⁵¹. Consequently, the classification of an AI system as prohibited or high-risk cannot be determined solely by the provider. Therefore, the provider should work in close cooperation with the deployer to fully understand the intended use of the system and to ensure appropriate compliance measures are taken.

²⁴⁸ AI Act, Annex III, paragraph 6: "(d) AI systems intended to be used by law enforcement authorities or on their behalf or by Union institutions, bodies, offices or agencies in support of law enforcement authorities for assessing the risk of a natural person offending or re-offending not solely on the basis of the profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680, or to assess personality traits and characteristics or past criminal behaviour of natural persons or groups". *Guidelines*, cited, paragraph 204.

²⁴⁹ The human intervention has been translated into the legal requirement of "human oversight" enshrined in Article 14 of the AI Act.

²⁵⁰ *Guidelines*, cited, paragraph 214.

²⁵¹ In similar terms MIRÓ LLINARES, SANTISTEBAN GALARZA, *The regulation of predictive policing systems in the Artificial Intelligence Act*, cited, 349.

High-risk classification also applies to systems used by, or on behalf of, law enforcement authorities to assess the likelihood that a natural person may become the victim of a criminal offence²⁵². While such systems may appear less concerning from a fundamental rights perspective – given that their purpose is to identify potential victims rather than potential offenders – they nonetheless involve the profiling of individuals.

Annex III, point 6 of the AI Act further identifies additional uses of AI systems that are classified as high-risk when deployed in the context of law enforcement. Specifically, the applications described under letters b), c), and e) concern AI systems employed during the investigative phase of criminal offences or in the subsequent prosecution process. These systems may include, for example, tools designed to detect, collect, and analyse evidence; systems supporting the identification of suspects; or applications assisting prosecutorial decision-making. The high-risk designation reflects the significant implications such systems may have for the fairness of criminal proceedings, the presumption of innocence, and the protection of fundamental rights. By intervening at crucial stages of investigation and prosecution, these AI tools influence not only the operational efficiency of law enforcement but also the legal outcomes and the rights of individuals subject to investigation or prosecution. The use of AI in criminal investigations and proceedings raises issues that lie beyond the scope of the present Thesis, inevitably invoking considerations of criminal procedural law and the procedural safeguards established therein²⁵³.

ANNEX III – HIGH-RISK AI SYSTEMS

PAR. 6, LETT. B)	AI systems intended to be used by or on behalf of law enforcement authorities or by Union institutions, bodies, offices or agencies in support of law enforcement authorities as polygraphs or similar tools;
PAR. 6, LETT. C)	AI systems intended to be used by or on behalf of law enforcement authorities, or by Union institutions, bodies, offices or agencies, in support of law enforcement authorities to evaluate the reliability of evidence in the course of the investigation or prosecution of criminal offences;
PAR. 6, LETT. E)	AI systems intended to be used by or on behalf of law enforcement authorities or by Union institutions, bodies, offices or agencies in support of law enforcement authorities for the profiling of natural persons as referred

²⁵² Annex III, par. 6, lett. a).

²⁵³ S. QUATTROCCOLO, *Artificial Intelligence, Computational Modelling and Criminal Proceedings. A Framework for A European Legal Discussion*, Springer, Legal Studies in International, European and Comparative Criminal Law, Cham, Switzerland, 2020. For the Italian debate on the topic see M. PISATI, *Indagini preliminari e intelligenza artificiale: efficienza e rischi per i diritti fondamentali*, in *Processo penale e giustizia*, 4, 2020, 957–971. L. PRESSACCO, *Intelligenza artificiale e ragionamento probatorio nel processo penale*, in *BioLaw Journal*, 4, 2022, 503–530. F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto penale e uomo*, 10(2), 2019. Available at: <https://air.unimi.it/handle/2434/680633>. G. DI PAOLO, L. PRESSACCO (Eds.), *Intelligenza Artificiale e Processo Penale. Indagini, Prove, Giudizio*, Università degli Studi di Trento, Quaderni Della Facoltà Di Giurisprudenza, 63, 2022.

to in Article 3(4) of Directive (EU) 2016/680 in the course of the detection, investigation or prosecution of criminal offences;

As far as place-based predictive policing systems are concerned they do not typically involve an assessment of individual risk. Unlike the Commission's proposal, which classified place-based predictive policing as high-risk, the final version, in line with the Council General Approach, neither prohibits such systems nor designates them as high-risk²⁵⁴. Rather, they fall within the category of minimal-risk AI systems and, as a consequence, are merely subject to non-binding voluntary Codes of Conduct²⁵⁵.

The exclusion of place-based predictive policing systems from both the prohibited and high-risk categories warrants careful scrutiny. Even where such systems do not entail the profiling of individuals, they nonetheless give rise to significant concerns regarding fundamental rights and freedoms, as examined in Chapter 2. As a result, this reclassification risks placing certain systems under a less stringent regulatory regime, characterised by reduced obligations and more limited oversight by market surveillance authorities. In this author's assessment, while the exclusion from the high-risk category may be justified for relatively simple systems – such as the Italian KeyCrime or Software XLAW – whose functionality is limited and whose outputs play a marginal or supportive role in decision-making, the same reasoning does not necessarily extend to more complex predictive policing systems. Systems that rely on multifactorial data inputs, advanced statistical or machine-learning techniques, and that exert a more direct influence on operational policing decisions may generate heightened risks for fundamental rights and freedoms. As the European Commission's Guidelines acknowledge, the distinction between person-based and place-based predictive policing systems is not always clear-cut. In some cases, for instance, a system may predict the locations where crimes are more likely to occur but then use the risk score of that location as a factor in profiling an individual. In such circumstances, the system could be regarded as person-based and thus fall under Article 5 or Article 6, depending on the specificities of the case²⁵⁶.

²⁵⁴ FERNANDO MIRÓ LLINARES, MARIO SANTISTEBAN GALARZA, *The regulation of predictive policing systems in the Artificial Intelligence Act*, cited, 347.

²⁵⁵ AI Act, Article 95.

²⁵⁶ *Guidelines*, cited, paragraph 213.

3. Automated video surveillance in the AI Act

In Chapter 2, I examined automated video surveillance (AVS), understood as the deployment of algorithmic tools to process and analyse video footage. AVS falls within the scope of the AI Act only insofar as AI systems are employed to enhance its performance. In such instances, an AI-enabled AVS does not, in itself, constitute a prohibited practice nor is it automatically classified as high-risk. Its legal characterization, however, shifts when the technology is combined with biometrics – such as in relation to face recognition technology (FRT) – or when it is deployed for law enforcement purposes, as outlined in Annex III, point 6. By contrast, other forms of AVS may fall within the minimal-risk category, and, as a result, only Codes of Conduct would apply. The subsequent section undertakes a detailed analysis of these two scenarios, namely the integration of AVS with biometric systems and its deployment for law enforcement purposes. Particular attention will be paid to the points at which these scenarios intersect, since such overlaps raise complex legal and practical questions regarding the scope of application of the AI Act and the adequacy of its safeguards for fundamental rights.

Building on the distinctions examined in Chapter 2, the AI Act differentiates between three core uses of biometrics: biometric identification, biometric verification, and biometric categorisation. Article 3(35) defines “biometric identification” as the automated recognition of physical, physiological, behavioural, or psychological characteristics of a natural person for the purpose of establishing his/her identity by comparing the individual’s biometric data with biometric data contained in a reference database. By contrast, “biometric verification”, as set out in Article 3(36), refers to the automated, one-to-one confirmation of a person’s identity by comparing his/her biometric data with data he/she has previously provided. Finally, while the AI Act does not provide a stand-alone definition of biometric categorisation, Article 3(40) defines “biometric categorisation AI systems” as those which assign natural persons to specific categories on the basis of their biometric data.

It is worth noting that the AI Act defines biometric data as “*means personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person [...]*”²⁵⁷. Unlike the GDPR and the LED, the AI Act does not refer to the “unique identification” of a natural person²⁵⁸. This omission raises an important point. Under the GDPR and the LED, classification *does* amount to biometric data processing when it relies on physical, physiological, or behavioural traits – such as facial geometry, fingerprints, voice patterns, gait, or keystroke dynamics –

²⁵⁷ AI Act, Article 3(34).

²⁵⁸ Despite this omission, Recital 14 of the AI Act stresses that “*the notion of ‘biometric data’ used in this Regulation should be interpreted in light of the notion of biometric data as defined in Article 4, point (14) of Regulation (EU) 2016/679, Article 3, point (18) of Regulation (EU) 2018/1725 and Article 3, point (13) of Directive (EU) 2016/680*”.

with the purpose or effect of uniquely identifying a natural person²⁵⁹; when biometric features are used exclusively for categorisation rather than identification – for example, systems that classify facial images by gender, ethnicity, or emotional state – the processing does not fall within the meaning of “biometric data” under Article 4(14) GDPR or Article 3(13) LED. By contrast, such categorisation AI systems *does* process biometric data within the meaning of the AI Act²⁶⁰. It may be argued that the AI Act introduces a broadened notion of “biometric data”, extending beyond its traditional focus on identification. Under this approach, biometric data also covers non-identifying uses, where biometric information is not employed to single out a specific individual but rather to extract features for categorising people, for instance by age, gender, or other traits²⁶¹. In any case, this distinction is crucial, since biometric categorisation or verification that does not aim at uniquely identifying an individual does not trigger the heightened safeguards afforded to biometric data under data protection legislation.

3.1. AI systems for categorising natural persons

The European legislator faced the task of determining whether the AI Act should regulate all forms of biometric recognition or only certain types, and how these technologies should be classified. In particular, the AI Act had to establish which practices ought to be prohibited, which should be treated as high-risk and subject to strict requirements, and which could fall outside these categories²⁶². With regard to AI systems used for categorisation, the legislative negotiations brought about significant changes. In both the Commission’s original proposal²⁶³ and the Council’s General Approach²⁶⁴, biometric categorisation systems were classified as limited-risk applications subject only to transparency obligations for deployers. By contrast, the European Parliament²⁶⁵ advocated for a prohibition of certain AI-based categorisation systems and the removal of the corresponding transparency obligations. In the final text of the AI Act, a compromise was reached²⁶⁶: categorisation systems are prohibited in certain cases, while in others they are classified as high-risk.

²⁵⁹ For a discussion on the definition of biometric data in the LED see L. TOSONI, L.A. BYGRAVE, *Article 3. Definitions*, in E. KOSTA, F. BOEHM (Eds.), *The Law Enforcement Directive: A Commentary*, Oxford University Press, 2024, 118-123 (*Definition of ‘biometric data’ (Article 3(13))*).

²⁶⁰ AI Act, Recital 14 “*Biometric data can allow for the authentication, identification or categorisation of natural persons and for the recognition of emotions of natural persons*”.

²⁶¹ See J. CZARNOCKI, *Will new definitions of emotion recognition and biometric data hamper the objectives of the proposed AIA?*, in International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 2021, 1-4.

²⁶² LEIRE ESCAJEDO SAN-EPIFANIO, *Biometric Recognition In The Artificial Intelligence Act: Exemptions, Prohibitions And High-Risk Specialties*, in LORENZO COTINO HUESO, DIANA-URANIA GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 183-184.

²⁶³ European Commission’s Proposal, cited.

²⁶⁴ Council’s General Approach, cited.

²⁶⁵ European Parliament’s Amendments, cited.

²⁶⁶ AI Act, cited.

CHAPTER 3

EUROPEAN COMMISSION'S PROPOSAL	COUNCIL'S APPROACH	GENERAL	EUROPEAN PARLIAMENT'S AMENDMENTS	FINAL VERSION
Article 1, lett. c); Article 52 – transparency obligations for biometric categorisation systems	Article 52 – transparency obligations for biometric categorisation systems		Amendment 217: Article 5 (b a) prohibits “the placing on the market, putting into service or use of biometric categorisation systems that categorise natural persons according to sensitive or protected attributes or characteristics or based on the inference of those attributes or characteristics. This prohibition shall not apply to AI systems intended to be used for approved therapeutical purposes on the basis of specific informed consent of the individuals that are exposed to them or, where applicable, of their legal guardian”	Article 5 (g) prohibits “the placing on the market, the putting into service for this specific purpose, or the use of biometric categorisation systems that categorise individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation; this prohibition does not cover any labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data or categorizing of biometric data in the area of law enforcement”
			Amendment 485 – Article 52(2): Users of an emotion recognition system or a biometric categorisation system which is not prohibited pursuant to Article 5 shall inform in a timely, clear and intelligible manner of the operation of the system the natural persons exposed thereto and obtain their consent prior to the processing of their biometric and other personal data in accordance with Regulation (EU) 2016/679, Regulation (EU) 2016/1725 and Directive (EU) 2016/280, as applicable. This obligation shall not apply to AI systems used for biometric categorisation, which are permitted by law to detect, prevent and investigate criminal offences	Article 50(3): Deployers of an emotion recognition system or a biometric categorisation system shall inform the natural persons exposed thereto of the operation of the system, and shall process the personal data in accordance with Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680, as applicable. This obligation shall not apply to AI systems used for biometric categorisation and emotion recognition, which are permitted by law to detect, prevent or investigate criminal offences, subject to appropriate safeguards for the rights and freedoms of third parties, and in accordance with Union law
Annex III, par. 1 Biometric identification and categorisation of natural persons: (a) AI systems intended to be used for the ‘real-time’ and ‘post’ remote biometric			Amendment 710 – Annex III, paragraph 1 reference to categorisation systems deleted	Annex III, paragraph 1, b) AI systems intended to be used for biometric categorisation, according to sensitive or protected attributes or

identification of natural persons;

characteristics based on the inference of those attributes or characteristics

Article 5(1)(g) of the AI Act prohibits AI systems used for categorising natural persons based on biometric data to deduce sensitive characteristics. For the prohibition to apply, five cumulative conditions must be satisfied²⁶⁷. First, the practice must amount to the placing on the market, the putting into service for this specific purpose, or the use of an AI system. Second, the system in question must qualify as a biometric categorisation system within the meaning of Article 3(40). Third, the AI system must be deployed for the purpose of categorising natural persons. Fourth, such categorisation must be carried out on the basis of the individual's biometric data. Finally, and most critically, the biometric data must be used *specifically* to deduce or infer one of the following sensitive attributes: race, political opinions, trade union membership, religious or philosophical beliefs, sex life, or sexual orientation. These characteristics mirror the special categories of personal data of Article 9 GDPR.

In 2012, the Article 29 Working Party released its Opinion²⁶⁸ where it defines categorisation by biometric systems as “*the process of establishing whether the biometric data of an individual belongs to a group with some pre-defined characteristic*”²⁶⁹. The purpose of a categorisation system is, by definition, to assign individuals to specific categories. The subsequent identification of persons who have already been categorised is merely a potential consequence of such categorisation, rather than its primary function.

The final version of the AI Act thus prohibits biometric categorisation only where it is based on sensitive characteristics, or on grounds protected under Union non-discrimination law²⁷⁰. Accordingly, a biometric categorisation system falls within the scope of the prohibition when it is used, for instance, to infer a person's race from their facial features. In certain instances, the categorisation of individuals on the basis of their biometric data, even where this involves the deduction of sensitive characteristics, may be necessary, for example, in order to assess whether the datasets used to train a model are sufficiently representative of the population in which the system will be deployed²⁷¹ or to carry out legitimate activities.

Accordingly, while the general rule in Article 5(1)(g) is that biometric categorisation systems are prohibited when they are used to infer sensitive characteristics, the European legislator identified two

²⁶⁷ *Guidelines*, cited, par. 273.

²⁶⁸ Article 29 Working Party, *Opinion 3/2012 on developments in biometric technologies*, WP193, 27 April 2012.

²⁶⁹ *Ibidem*, 6.

²⁷⁰ For a critical analysis of this provision see LEIRE ESCAJEDO SAN-EPIFANIO, *Biometric Recognition In The Artificial Intelligence Act: Exemptions, Prohibitions And High-Risk Specialties*, cited, 216-217.

²⁷¹ In such cases, categorisation may serve a corrective purpose, namely to prevent discriminatory outcomes. *Guidelines*, cited, paragraph 285.

broad, ambiguous exceptions. The lack of clarity in delineating the two exceptions has been addressed through interpretative guidance, most notably in Recital 30 of the AI Act and in the European Commission’s Guidelines, which together contribute to resolving the uncertainty. First, Article 5(1)(g) explicitly excludes from the prohibition “*any labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data*”. For example, categorisation AI systems may be used to check whether a training dataset contains enough diversity (e.g. to avoid a face recognition model working better on light-skinned men than on dark-skinned women) and equally represent all demographic groups. In this case, the system may tag (“label”) or sort (“filter”) images based on biometric features, but the goal is to ensure fairness and quality of the dataset, not to deploy categorisation against people in real-world use. For the practice to fall outside the scope of the prohibition, the datasets must have been acquired lawfully, that is, in accordance with data protection legislation and on the basis of a specific legal ground. Second, a similar exception applies where biometric categorisation systems are used to label or filter lawfully acquired datasets in the area of law enforcement. Law enforcement authorities may thus categorise biometric data for their own legitimate activities, such as for sorting lawfully collected images in a police database.

While Article 5(1)(g) prohibits the use of biometric categorisation AI systems that infer the sensitive characteristics explicitly listed therein, Annex III point 1, letter b) classifies as high-risk those biometric categorisation systems that target other protected attributes, such as age, disability, gender, or similar traits recognised under equality law but not expressly covered by Article 5(1)(g)²⁷². Therefore, while the original draft classified biometric categorisation AI systems as limited-risk, subject only to transparency obligations, the final version, aligned with the Parliament’s position, provides stronger safeguards for the fundamental rights and freedoms of individuals.

As a general rule, the AI Act requires deployers of biometric categorisation systems to inform the individuals subject to their operation. Some scholars argue that it remains unclear what this provision contributes beyond the safeguards already enshrined in data protection legislation, particularly Article 13 of the GDPR. They critically hypothesize that its purpose may lie in addressing situations where emotion recognition or biometric classification systems do not process personal data. Yet, they contend that such an interpretation would entail a contentious and overly restrictive reading of the GDPR, given that the operation of these systems almost invariably involves the processing of personal data²⁷³. In

²⁷² AI Act, Annex III, paragraph 1, letter b) considers high-risk those “*AI systems intended to be used for biometric categorisation, according to sensitive or protected attributes or characteristics based on the inference of those attributes or characteristics*”.

²⁷³ M. VEALE, F. ZUIDERVEEN BORGESIU, *Demystifying the Draft EU Artificial Intelligence Act*, cited, 107. It reflects the ongoing debate over whether emotional recognition and biometric classification systems necessarily process personal data under the GDPR. Some developers, such as the Fraunhofer Institute with its *Anonymous Video Analytics for Retail and Digital Signage* (AVARD) system, argue that such technologies only generate anonymous, aggregate outputs and therefore

any case, this obligation does not apply where such systems are lawfully used to detect, prevent, or investigate criminal offences, provided that appropriate safeguards for the rights and freedoms of third parties are in place and that the use is consistent with Union law²⁷⁴.

3.2. Biometric identification systems

In relation to biometric identification systems, the AI Act establishes a complex regulatory framework. It is particularly noteworthy that, in drafting these provisions, the European legislator appears to have had face recognition technology foremost in mind.

EUROPEAN COMMISSION'S PROPOSAL ²⁷⁵	COUNCIL'S GENERAL APPROACH ²⁷⁶	EUROPEAN PARLIAMENT'S AMENDMENTS ²⁷⁷	FINAL VERSION ²⁷⁸
Article 5, letter d) prohibits “the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement, unless and in as far as such use is strictly necessary for one of the following objectives: (i) the targeted search for specific potential victims of crime, including missing children; (ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack; (iii) the detection, localisation, identification or prosecution of a perpetrator or suspect of a criminal offence referred to	Article 5, letter d) prohibits “the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces by law enforcement authorities or on their behalf for the purpose of law enforcement, unless and in as far as such use is strictly necessary for one of the following objectives: (i) the targeted search for specific potential victims of crime; (ii) the prevention of a specific and substantial threat to the critical infrastructure, life, health or physical safety of natural persons or the prevention of terrorist attacks;	Article 5, letter d) prohibits “the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces”;	Article 5 letter (h) prohibits “the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement, unless and in so far as such use is strictly necessary for one of the following objectives: (i) the targeted search for specific victims of abduction, trafficking in human beings or sexual exploitation of human beings, as well as the search for missing persons; (ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or a genuine and present or genuine and foreseeable threat of a terrorist attack;

do not fall within the GDPR's scope. However, this interpretation is highly contentious, as many Data Protection Authorities, national courts, and legal scholars maintain that personal data is inevitably processed in these situations, since biometric or emotional features of identifiable individuals are captured and analysed, even if the end result is anonymised. Accepting the narrower view risks implicitly legitimising a restrictive reading of the GDPR, weakening protections by allowing developers to bypass data protection obligations through claims of anonymisation.

²⁷⁴ AI Act, Article 50(3).

²⁷⁵ European Commission's Proposal, cited. For a critical analysis of the prohibition set out in the proposal see I. BARKANE, *Questioning the EU Proposal for an Artificial Intelligence Act: The Need for Prohibitions and a Stricter Approach to Biometric Surveillance*, in *Information Polity*, 27(2), 2022, 147-162.

²⁷⁶ Council's General Approach, cited.

²⁷⁷ European Parliament's Amendments, cited.

²⁷⁸ AI Act, cited.

CHAPTER 3

in Article 2(2) of Council Framework Decision 2002/584/JHA⁶² and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years, as determined by the law of that Member State

(iii) the localisation or identification of a natural person for the purposes of conducting a criminal investigation, prosecution or executing a criminal penalty for offences, referred to in Article 2(2) of Council Framework Decision 2002/584/JHA³² and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years, or other specific offences punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least five years, as determined by the law of that Member State.

(iii) the localisation or identification of a person suspected of having committed a criminal offence, for the purpose of conducting a criminal investigation or prosecution or executing a criminal penalty for offences referred to in Annex II and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least four years. Point (h) of the first subparagraph is without prejudice to Article 9 of Regulation (EU) 2016/679 for the processing of biometric data for purposes other than law enforcement.

Article 5, letter (d) prohibits “the putting into service or use of AI systems for the analysis of recorded footage of publicly accessible spaces through ‘post’ remote biometric identification systems, unless they are subject to a pre-judicial authorisation in accordance with Union law and strictly necessary for the targeted search connected to a specific serious criminal offense as defined in Article 83(1) of TFEU that already took place for the purpose of law enforcement.

Annex III, paragraph 1: Biometric identification and categorisation of natural persons:
(a) AI systems intended to be used for the ‘real-time’ and ‘post’ remote biometric identification of natural persons;

Annex III, paragraph 1 Biometrics:
(a) Remote biometric identification systems.

Annex III, paragraph 1, letter (a)
AI systems intended to be used for biometric identification of natural persons, with the exception of those mentioned in Article 5;

Annex III, paragraph 1. Biometrics, in so far as their use is permitted under relevant Union or national law:
(a) remote biometric identification systems. This shall not include AI systems intended to be used for biometric verification the sole purpose of which is to confirm that a specific natural person is the person he or she claims to be;

The final AI Act diverges considerably from the Parliament's position, which would have banned biometric identification systems outright, regardless of the sector of deployment.²⁷⁹ Closely aligned with the original proposal of the Commission and the Council's General Approach, the final version of Article 5(1)(h) expressly prohibits the use of real-time remote biometric identification systems for law enforcement purposes, subjecting it to certain exceptions²⁸⁰.

To fall within the scope of the prohibition, all of the following conditions must be satisfied cumulatively. First, the AI system must be a real-time remote biometric identification (hereinafter "RBI") system as defined by Articles 3(41) and 3(42) of the AI Act²⁸¹. Second, the practice must consist in the "use" of such system. Therefore, the placing on the market and the putting into service of RBI systems are allowed²⁸². Third, it must be used in publicly accessible spaces. Fourth, it must be used for law enforcement purposes, as defined in Chapter 1. As noted in Chapter 1, the AI Act does not apply to national security, military, or defence purposes, despite these representing the primary contexts for RBI deployment²⁸³.

A primary distinction relevant to the prohibition's scope concerns remote versus non-remote biometric identification systems. The difference lies in whether the natural person is actively involved in the capture of his or her biometric data. In the case of remote systems, the data are collected at a distance, without the subject's direct participation or even awareness²⁸⁴. An example of a RBI system is the use of cameras installed along public streets that automatically scan and analyse the faces of passers-by in order to identify individuals for surveillance purposes²⁸⁵. Similarly, body-worn cameras equipped with AI systems, when used by law enforcement officers during public demonstrations, are also regarded

²⁷⁹ Already in 2021 the European Parliament called for a restrictive position on biometric recognition. See European Parliament, *Resolution of 6 October 2021 on Artificial Intelligence in criminal law and its use by law enforcement authorities in criminal matters* (2020/2016(INI)), cited.

²⁸⁰ AI Act, Recital 32: "*The use of AI systems for 'real-time' remote biometric identification of natural persons in publicly accessible spaces for the purpose of law enforcement is particularly intrusive to the rights and freedoms of the concerned persons, to the extent that it may affect the private life of a large part of the population, evoke a feeling of constant surveillance and indirectly dissuade the exercise of the freedom of assembly and other fundamental rights. Technical inaccuracies of AI systems intended for the remote biometric identification of natural persons can lead to biased results and entail discriminatory effects*".

²⁸¹ AI Act, Article 3(41) "'remote biometric identification system' means an AI system for the purpose of identifying natural persons, without their active involvement, typically at a distance through the comparison of a person's biometric data with the biometric data contained in a reference database"; Article 3(42) "'real-time remote biometric identification system' means a remote biometric identification system, whereby the capturing of biometric data, the comparison and the identification all occur without a significant delay, comprising not only instant identification, but also limited short delays in order to avoid circumvention".

²⁸² This is justified by the fact that RBI systems used for purposes other than law enforcement may still be deployed, provided that they comply with the requirements applicable to high-risk AI systems

²⁸³ AI Act, Article 2(3).

²⁸⁴ AI Act, Recital 17.

²⁸⁵ *Guidelines*, cited, paragraph 306.

as fulfilling the criterion of remoteness in biometric identification. By contrast, non-remote systems generally presuppose the individual's active involvement, for example by presenting their face or fingerprint to a scanner²⁸⁶. This distinction also makes it clear why AI systems designed for biometric verification (*i.e.*, one-to-one authentication) fall outside the scope of "remote" identification: verification typically requires the person's deliberate and conscious participation²⁸⁷.

While there is broad consensus that *remote* biometric identification systems constitute the most intrusive and dangerous form of biometric technology, in my assessment, it is questionable that *non-remote* biometric identification systems are excluded even from the high-risk category, despite their potential negative implications for fundamental rights and freedoms. This is particularly evident when the following considerations are examined. First, the AI Act does not provide a clear definition of what constitutes "active participation"²⁸⁸. This creates legal uncertainty, as it is not evident whether "active participation" should be interpreted narrowly, requiring, for instance, a deliberate physical action by the individual such as placing a finger on a scanner, or more broadly, encompassing situations where the individual is aware of and consents to the biometric operation, even if the action itself is minimal. Second, non-remote biometric identification systems involve the processing of biometric data. Such data is classified as a "special category" under both the LED and the GDPR, thereby triggering the application of particularly stringent limitations and safeguards under EU data protection law²⁸⁹.

A second, equally relevant, distinction is that between *real-time* and *post* remote biometric identification systems. Real-time systems process biometric data instantaneously, near-instantaneously or in any event without a significant delay, allowing for immediate identification of individuals in live settings²⁹⁰. Post biometric identification systems, on the other hand, operate on pre-recorded material, such as recordings from CCTV cameras, processing biometric data and identify the natural person after a significant delay. The distinction between the two forms of biometric identification is not always clear-cut and must be assessed on a case-by-case basis²⁹¹. Despite this uncertainty, the distinction remains central to the AI Act, as a system's specific characteristics determine the risk category to which it is assigned. Therefore, the distinction rests solely on an uncertain temporal criterion²⁹², yet generates

²⁸⁶ AI Act, Recital 15.

²⁸⁷ "That exclusion is justified by the fact that such systems are likely to have a minor impact on fundamental rights of natural persons compared to the remote biometric identification systems which may be used for the processing of the biometric data of a large number of persons without their active involvement". AI Act, Recital 15.

²⁸⁸ ESCAJEDO SAN-EPIFANIO, *Biometric Recognition In The Artificial Intelligence Act: Exemptions, Prohibitions And High-Risk Specialties*, cited, 207.

²⁸⁹ ESCAJEDO SAN-EPIFANIO, *Biometric Recognition In The Artificial Intelligence Act: Exemptions, Prohibitions And High-Risk Specialties*, cited, 207-208.

²⁹⁰ AI Act, Recital 17.

²⁹¹ *Guidelines*, cited, paragraph 310.

²⁹² *Guidelines*, cited, paragraph 311.

fundamentally different legal consequences: real-time RBI systems are prohibited, whereas post RBI systems are permitted subject to high-risk requirements.

This sharp regulatory divergence, predicated solely on temporal distinctions, risks appearing arbitrary and undermining the coherence of the risk-based framework, raising questions about whether the categorization adequately captures the substantive risks posed by both practices²⁹³. On the one hand, the different treatment of real-time and post-biometric identification systems is justified by the fact that, in the case of real-time RBI systems, there are limited opportunities for audits, revisions, checks or corrections, and therefore mistakes may be more easily made²⁹⁴. On the other hand, it is difficult, in my analysis, to justify the assertion that identifying individuals, for example, at protests after the event is less intrusive than immediate identification²⁹⁵. Given the particular risks these systems pose to fundamental rights, the AI Act establishes specific regulatory obligations governing their deployment, while permitting Member States to adopt more restrictive measures²⁹⁶.

The regulatory framework permits deployment of these systems beyond law enforcement contexts, including by private commercial actors. Critically, no prohibition applies to highly sensitive areas such as migration management and border control, despite the notoriously opaque distinction between administrative and law enforcement activities in these domains. This omission raises profound concerns about potential systemic abuse²⁹⁷. Another aspect to be taken into account is that the AI Act prohibits only the *use* of real-time RBI systems, without extending this restriction to their placing on the market or their putting into service. This regulatory asymmetry gives rise to a paradox: while EU vendors remain entitled to commercialise and export such systems to jurisdictions where their use is authorised, their deployment within the Union is prohibited, or, more accurately, subject to narrowly circumscribed exceptions²⁹⁸. This discrepancy inevitably raises questions concerning regulatory coherence, the externalisation of ethical risks, and the European Union's capacity to exert normative influence in shaping global standards for biometric technologies²⁹⁹.

²⁹³ This choice was criticised in EDPB-EDPS, *Joint Opinion 5/2021*, cited, 11.

²⁹⁴ AI Act, Recital 32.

²⁹⁵ BARKANE, *Questioning the EU proposal for an artificial intelligence Act: The need for prohibitions and a stricter approach to biometric surveillance*, cited.

²⁹⁶ AI Act, Article 26(10).

²⁹⁷ Amnesty International, *EU: Artificial intelligence rulebook fails to stop proliferation of abusive technologies*, cited; L.L. CABRERA, *EU AI act brief - Pt. 2, privacy & surveillance*, cited.

²⁹⁸ In these terms see VEALE, ZUIDERVEEN BORGESIU, *Demystifying the draft EU artificial intelligence act-analysing the good, the bad, and the unclear elements of the proposed approach*, cited. Amnesty International presented some examples of these practices. Among them, the French firm Idemia/Morpho selling facial recognition to the Shanghai Public Security Bureau, or the Dutch firm Noldus selling facial expression analysis tool "FaceReader" to the Chinese Ministry of Public Security. Amnesty International, *Out of Control: Failing EU Laws for Digital Surveillance Export*, September 2020. Available at: <https://perma.cc/2GU5-84ZT>.

²⁹⁹ Some authors correctly stress that "one can question whether the EU institutions grant all people the same level of protection of human and fundamental rights or if the EU enforces a two-tier system in which different standards apply to

Although the use of such systems is generally prohibited, there may be cases where the use of real-time RBI systems is necessary and where the security benefits they offer outweigh their negative impact on fundamental rights and freedoms³⁰⁰. Article 5 thus identifies three exceptions³⁰¹. First, real-time RBI systems may be admissible for the targeted search of victims of three specific serious crimes and missing persons³⁰². Second, such systems may be used for preventive purposes in order to prevent a specific, substantial and imminent threat to the life or physical safety of natural persons or a genuine and present or genuine and foreseeable threat of a terrorist attack³⁰³. Third, real-time RBI systems may be used for the localisation or identification of suspects and offenders of certain serious crimes listed in Annex II of the AI Act³⁰⁴.

The first exception is relatively straightforward, as it expressly refers to three categories of serious criminal offences, namely abduction, trafficking in human beings and sexual exploitation of human beings. In addition, it authorises the use of real-time RBI systems for the search of missing persons³⁰⁵. The use of RBI systems in such situations is deemed necessary in order to protect victims of such crimes. The second exception is more complex, as it concerns preventive activities and leaves significant room for interpretation. Divergences among Member States are therefore likely to occur. As the Guidelines emphasise, the notion of an “*imminent threat to life or the physical safety of natural persons*” is ultimately determined at the national level, in accordance with domestic legislation, which must specify the conditions under which the use of real-time RBI systems may be authorised. Likewise, the definition of a “terrorist attack” is established by each Member State and may vary considerably across the Union. It should be underlined, however, that this preventive use of RBI is strictly confined to the most serious circumstances, namely, terrorist attacks or imminent threats to life or the physical safety of individuals. Notably, the provision does not make any reference to the broader notion of “*public security*”. Therefore, it cannot be used a real-time face recognition system in public street

EU and non-EU citizens”. RUDSCHIES, SCHNEIDER, *The Long and Winding Road to Bans for Artificial Intelligence: From Public Pressure and Regulatory Initiatives to the EU AI Act*, cited, 14.

³⁰⁰ AI Act, Recital 33 “*The use of those systems for the purpose of law enforcement should therefore be prohibited, except in exhaustively listed and narrowly defined situations, where the use is strictly necessary to achieve a substantial public interest, the importance of which outweighs the risks*”.

³⁰¹ A. GIANNINI, S. TAS, *AI Act and the Prohibition of Real-Time Biometric Identification. Much ado about nothing?*, 10 December 2024. Available at: <https://verfassungsblog.de/ai-act-and-the-prohibition-of-real-time-biometric-identification/>.

³⁰² AI Act, Article 5(1)(h)(i).

³⁰³ AI Act, Article 5(1)(h)(ii).

³⁰⁴ AI Act, Article 5(1)(h)(iii). These crimes are: terrorism, trafficking in human beings, sexual exploitation of children, and child pornography, illicit trafficking in narcotic drugs or psychotropic substances, illicit trafficking in weapons, munitions or explosives; murder, grievous bodily injury; illicit trade in human organs or tissue; illicit trafficking in nuclear or radioactive materials; kidnapping, illegal restraint or hostage-taking; crimes within the jurisdiction of the International Criminal Court; unlawful seizure of aircraft or ships; rape; environmental crime; organised or armed robbery; sabotage; participation in a criminal organisation involved in one or more of the offences listed above. Some of the offences listed in Annex II AI Act are the so-called ‘euro crimes listed in Article 83 TFEU. The list of criminal offences provided in Annex II is based on the 32 criminal offences listed in the Council Framework Decision 2002/584/JHA.

³⁰⁵ *Guidelines*, cited, 9.3.2.

based on general security or crime prevention concerns. Consequently, the use of RBI systems for the protection of public security is admissible only where such situations amount to a terrorist attack or an imminent threat to life or physical safety. It is thus the seriousness of the potential offences that justifies the admissibility of real-time RBI systems³⁰⁶. The third exception is the one most likely to give rise to divergences. This is because Article 5(1)(h)(iii) allows the deployment of real-time RBI systems for the investigation – when the crime has been committed and the perpetrator must be found – or the prosecution or execution of a criminal penalty – when the perpetrator has been found – of offences listed in Annex II, provided that such offences are punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least four years. The absence of a clear definition of such offences, combined with the diverse sensibilities of Member States where certain crimes may carry harsher penalties in some jurisdictions than in others creates the potential for significant divergences in application across the Union³⁰⁷.

3.2.1. Exceptions to the prohibition: safeguards

It has been affirmed that the exceptions outlined above risk undermining the substance of the prohibition. To mitigate this and to ensure that real-time RBI systems admitted under these exceptions operate in full compliance with fundamental rights and freedoms, the European legislator has introduced, in Articles 5(2) to 5(7), a series of safeguards and conditions.

The first of these stipulates that real-time RBI systems may be used only for the purpose of “*confirming the identity*” of a specifically targeted individual. This requirement is intended to prevent the deployment of these systems for indiscriminate or large-scale biometric surveillance of the general population. As clarified in the Guidelines, this safeguard embodies a balancing exercise: on the one hand, it weighs the seriousness of the situation that necessitates the use of a real-time RBI system and the potential harm that could result from refraining from its use; on the other hand, it measures these considerations against the impact that such use has on the rights and freedoms of the individual concerned. Since the use of such AI systems is limited to confirming the identity of an individual by comparing data collected in real time with data contained in a reference database, it follows that law enforcement authorities must already have grounds to believe, or reliable information indicating, that the targeted individual is either a victim of the crimes listed in Article 5(1)(h)(i) AI Act, or a suspect

³⁰⁶ *Guidelines*, cited, paragraph 347.

³⁰⁷ In relation to rape, for example, see the in-depth analysis I. ZAMFIR, *Definitions of rape in the legislation of EU Member States*, European Parliamentary Research Service, May 2025, PE 757.618.

or perpetrator in the situations envisaged in Article 5(1)(h)(ii) and (iii)³⁰⁸. Article 5(2) indicates the need to take into account a number of aspects when deploying a real-time RBI system, namely the nature of the situation giving rise to the possible use and, in particular, the seriousness³⁰⁹, probability³¹⁰ and scale³¹¹ of the harm for natural persons, society and law enforcement purposes that would be caused if the system were not used. At the same time, law enforcement authorities must consider the consequences of the use of the system on the rights and freedoms of targeted individuals, in particular, the seriousness, probability and scale of those consequences. This assessment should become part of the Fundamental Rights Impact Assessment (hereinafter “FRIA”), requested by Article 5(2), which must comply with the conditions laid down in Article 27 AI Act³¹². In addition, the RBI system must be registered in the EU database provided for in Article 49 AI Act³¹³. However, since law enforcement purposes must be safeguarded, Article 49(4) requires that the relevant information be recorded in a secure, non-public section of the database, containing only limited information and subject to strictly restricted access (see Chapter 4, Part I)³¹⁴.

Another condition is that the use of RBI systems must be expressly authorised by a legal provision within the national legal framework³¹⁵. Any such provision must clearly specify the temporal, geographical and personal limitations of the system’s deployment, and must comply with the conditions set out in Article 5(5). In practice, this means that the use of an RBI system may only be authorised for a limited period of time, strictly necessary to achieve the specific purpose for which it is exempted; it must be confined to one or more defined areas, identified in a non-discriminatory manner; and it must also set out the categories of persons who may be targeted by the system.

The added value of this provision is not particularly striking. Even without the explicit reference to the necessity of a legal basis authorising the use of such systems, it should already have been evident, given that Member States operate under constitutional systems grounded in the rule of law, that such authorisation is required³¹⁶. Furthermore, data protection legislation has already required a legal basis

³⁰⁸ *Guidelines*, cited, paragraph 259

³⁰⁹ The degree of interference with the fundamental rights at stake.

³¹⁰ The number and categories of persons affected by the interference.

³¹¹ The likelihood that an event will occur.

³¹² FRIA should complement the DPIA performed by the deployer as data controller, without covering aspects already addressed in the DPIA and avoiding overlaps. For an analysis see A. MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template*, in *Computer Law & Security Review*, 54, 2024.

³¹³ However, in the case of a duly justified emergency (such as an imminent threat), deployment may start even prior to registration, provided the law enforcement authority registers the system without undue delay.

³¹⁴ AI Act, Article 49(4).

³¹⁵ *Guidelines*, cited, paragraphs 365-368.

³¹⁶ This has been emphasised by ESCAJEDO SAN-EPIFANIO, *Biometric Recognition In The Artificial Intelligence Act: Exemptions, Prohibitions And High-Risk Specialties*, cited, 222 and 225. It is important to note that the AI Act does not in itself provide a legal basis for data processing, with the sole exception of Article 10(5), justify restrictions on fundamental rights, or satisfy requirements of criminal, sanctioning, or procedural legality.

for using systems that process biometric data³¹⁷. In any case, although this provision does not introduce anything substantially new, it is nevertheless useful to recall that, should Member States decide to authorise such uses, they must adopt a detailed legal measure and notify it to the Commission within thirty days of its adoption.

Before an RBI system can be used, Article 5(3) of the AI Act requires a prior authorisation for each individual use³¹⁸. Such authorisation must be granted by a judicial authority or an independent administrative authority, whose decision is binding in the Member State where the system is to be deployed. This mechanism ensures a prior assessment of whether the use of the RBI system is necessary and proportionate to the purpose pursued³¹⁹, and strictly limited in terms of duration, geographic scope, and categories of persons concerned, as defined in the legal basis. Therefore, while the deployer is responsible for carrying out an initial assessment documented in the FRIA, a second, independent assessment is performed by the judicial or administrative authority.

The concept of “*independency*” is crucial, since the authority that grant the authorisation need to be independent from the authority that uses the RBI system. In a constitutional system guided by the rule of law, judicial authorities are generally independent³²⁰. The CJEU has clarified that, as regards public bodies, independence usually refers to a status that guarantees the ability to act completely freely in relation to those entities from which independence must be safeguarded – in this case, the law enforcement authority using the real-time RBI system – remaining shielded from any instructions or pressure³²¹. If the authority were to act under executive orders, its decisions could be shaped by political or security-driven considerations rather than by the objective application of the law and the protection of individual rights. Safeguarding institutional independence is thus a prerequisite to ensuring effective oversight, impartial enforcement, and public trust in the governance of AI systems used for biometric identification. Authorisation shall only be granted upon a reasoned, substantiated and motivated request from the competent law enforcement authority. This request must clearly specify the intended use and the conditions under which it will be carried out. It must also take into account all the aspects outlined above and be in accordance with the detailed rules of national law referred to

³¹⁷ The EDPB clarified that a mere transposition may not be invoked as a legal basis for the processing of biometric data, as it would lack both precision and foreseeability. It is thus essential to have a specific provision providing for such a legal basis in the national law.

³¹⁸ *Guidelines*, cited, paragraph 388.

³¹⁹ Depending on the situation, all the elements must be “justified”.

³²⁰ Judicial independence is one of the crucial facets of the rule of law and is guaranteed by Article 47 of the CFREU and Article 6(1) of the ECHR. An authority involved in previous investigations, in that case the public prosecutor, does not have such independence.

³²¹ CJEU, Judgement of 13 June 2018, *Commission v Poland*, C-530/16, ECLI:EU:C:2018:430, paragraph 67.

in paragraph 5³²². Furthermore, authorisation for the use of the real-time RBI system in publicly accessible spaces must only be granted if the relevant law enforcement authority has completed a FRIA and registered the system in the database, as requested by Article 5(2).

As a rule, the authorisation shall be granted *before* the use of the real-time RBI system. The use of the real-time RBI system is admissible in duly justified situations on grounds of urgency without prior authorisation, provided it can be requested without undue delay and at the latest within twenty-four hours³²³. Law enforcement authorities must also notify the relevant market surveillance authority and the national Data Protection Authority of each use of the real-time RBI system, which will then submit an annual report on the use of such systems to the Commission. In any case, when authorisation is granted for the deployment of a real-time RBI system, the deployer shall not make any decisions that have adverse legal consequences for a person based solely on the system's output³²⁴. Even if it is not specifically stated in Article 5, exempted AI biometric identification systems are considered high-risk and therefore they need to respect all the provisions related to such category³²⁵.

Therefore, if a Member State opts to fully or partially permit the use of real-time RBI systems in publicly accessible spaces for law-enforcement purposes – within the limits and under the conditions set out in Article 5(1)(h) – it must establish, through national legislation, a comprehensive framework governing such authorisations. This framework must lay down detailed rules on how authorisations are to be requested, issued, exercised, supervised, and reported. It must also clearly specify which of the objectives listed in Article 5(1)(h), including which of the criminal offences referred to in point (h)(iii), may justify the use of these systems by competent authorities for law enforcement purposes. Using SARI Real-Time as an illustrative example, if Italy decides to authorise the deployment of such a system – falling within the definition of real-time RBI – national law must provide a specific regulatory basis encompassing all the requirements described above. In this regard, Italy's Law No. 132/2025 has empowered the Government to adopt detailed regulations governing the use of AI

³²² AI Act, Article 5(5). “A Member State may decide to provide for the possibility to fully or partially authorise the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement within the limits and under the conditions listed in paragraph 1, first subparagraph, point (h), and paragraphs 2 and 3. Member States concerned shall lay down in their national law the necessary detailed rules for the request, issuance and exercise of, as well as supervision and reporting relating to, the authorisations referred to in paragraph 3. Those rules shall also specify in respect of which of the objectives listed in paragraph 1, first subparagraph, point (h), including which of the criminal offences referred to in point (h)(iii) thereof, the competent authorities may be authorised to use those systems for the purposes of law enforcement. Member States shall notify those rules to the Commission at the latest 30 days following the adoption thereof. Member States may introduce, in accordance with Union law, more restrictive laws on the use of remote biometric identification systems”.

³²³ If the authorisation is rejected, the use of real-time biometric identification systems linked to that authorisation should cease with immediate effect and all the data related to such use should be discarded and deleted.

³²⁴ In line with Article 14 of the AI Act.

³²⁵ This can be understood by the *Guidelines*, cited, par. 37: “Conversely, most AI systems that fall under an exception from a prohibition listed in Article 5 AI Act will qualify as high-risk”. In the same terms, see A. GIANNINI, S. TAS, *AI Act and the Prohibition of Real-Time Biometric Identification. Much ado about nothing?*, cited.

systems in the law enforcement context, thereby creating the legal space necessary to introduce rules specifically addressing real-time RBI systems.

3.2.2. High-risk biometric identification systems

RBI systems that are permitted under relevant Union or national law are classified within the high-risk category under the AI Act. Annex III, point 1, provides a general reference to biometric identification systems, thereby establishing that, irrespective of the operational scenario in which they are deployed, such systems raise significant risks for fundamental rights and thus warrant stricter regulatory control. In addition, Annex III, point 6, addresses operational scenarios in the context of law enforcement. While it does not explicitly mention biometric recognition, some of the scenarios it outlines may indirectly encompass potential uses of biometric technologies. For instance, this could include biometric systems used to assess the risk of criminality or recidivism (point 6(d)), or to evaluate the reliability of a witness or expert witness as part of the broader assessment of evidentiary reliability (point 6(c))³²⁶.

In both instances, the classification of biometric recognition systems as high-risk systems entails the application of the strict requirements and obligations established in Chapter III, Sections 2 and 3 of the AI Act. Beyond these general requirements, the AI Act also introduces specific provisions that apply exclusively to (non-prohibited) biometric identification systems. First, in the context of an investigation involving the targeted search for a person suspected or convicted of a criminal offence, a law enforcement authority intending to deploy an AI system for post-remote biometric identification must obtain prior authorisation from a judicial authority, or from an administrative authority whose decision is binding and subject to judicial review. However, this requirement may be waived when the system is used solely for the initial identification of a potential suspect, provided that the identification is based on objective and verifiable facts directly linked to the offence³²⁷.

Unlike exempted uses of biometric identification systems, post-remote biometric identification does not require that the authorisation be granted by an *independent* administrative authority. The rationale for this distinction is not entirely clear. In this author's assessment, a plausible explanation is that the requirement of independence under Article 5 reflects the exceptional nature of exemptions from a

³²⁶ In these terms see ESCAJEDO SAN-EPIFANIO, *Biometric Recognition In The Artificial Intelligence Act: Exemptions, Prohibitions And High-Risk Specialties*, cited, 228.

³²⁷ AI Act, Article 26. In any case, if the authorisation requested pursuant to the first subparagraph is rejected, the use of the post-remote biometric identification system linked to that requested authorisation shall be stopped with immediate effect and the personal data linked to the use of the high-risk AI system for which the authorisation was requested shall be deleted.

prohibition, whereas in the case of high-risk but not prohibited uses (such as post-remote identification), the AI Act allows authorisation by an administrative authority, provided its decision is binding and subject to judicial review. This approach appears to rest on the assumption that post-remote identification is less intrusive than real-time mass surveillance, and in such cases judicial review is intended to serve as the safeguard against potential misuse. In this author's assessment, such a distinction is not completely justifiable, as post-RBI systems, although less intrusive than real-time ones, nonetheless infringe upon individuals' rights and freedoms, as discussed in Chapter 2. In any case, significant uncertainty remains as to how this mechanism will operate in practice, particularly regarding whether the authorising authority could, in effect, be the same law enforcement body seeking to deploy the system.

Furthermore, the use of post RBI systems for law enforcement purposes shall not be carried out in an untargeted manner, that is, *“without any link to a criminal offence, a criminal proceeding, a genuine and present or genuine and foreseeable threat of a criminal offence, or the search for a specific missing person”*³²⁸. In addition, human intervention is always required and law enforcement authorities may not take decisions that produce adverse legal effects on a person based solely on the output of such systems³²⁹. The same provision states that the use of such high-risk AI systems shall be documented in the relevant police file and shall be made available to the relevant market surveillance authority and the national Data Protection Authority upon their request³³⁰. Law enforcement authorities shall submit annual reports to the relevant market surveillance and national Data Protection Authorities on their use of post-remote biometric identification systems³³¹.

3.3. Biometric verification

In the final version of the AI Act, AI systems that are intended to be used for biometric verification, including authorisation purposes, fall outside the scope of the prohibition in Article 5(1)(h) AI Act and of the high-risk category. While the Commission's proposal made no reference to biometric verification, the Council's General Approach explicitly began to consider biometric verification systems, albeit only in the context of their exclusion from the notion of biometric identification systems. It was through the Parliament's amendments that a formal definition of biometric verification

³²⁸ AI Act, Article 26(10).

³²⁹ This is in line with Article 14 of the AI Act.

³³⁰ It is excluded the disclosure of sensitive operational data related to law enforcement. In addition to this duty, the powers conferred by Directive (EU) 2016/680 on supervisory authorities remain unaffected.

³³¹ It is excluded the disclosure of sensitive operational data related to law enforcement.

was introduced into the Regulation³³². What distinguishes biometric verification from biometric identification is mainly the absence of a reference database: in biometric identification, an individual's biometric data (the “one”) is compared against a database containing the biometric data of many others (the “many”), whereas biometric verification does not rely on such a database³³³.

The underlying rationale is that biometric verification systems are generally regarded as posing a much lower risk³³⁴. This is because they are typically applied to a specific individual – not a large number of persons – require the active participation of the person concerned and are not deployed for surveillance purposes. In this author's assessment, however, it is debatable that biometric verification systems are completely excluded from the high-risk category under the AI Act. In fact, although biometric verification is generally less intrusive than biometric identification, since it focuses on specific individuals and usually requires their active participation, such systems are not without risks. For example, in law enforcement contexts, face verification may be used to confirm a suspect's identity before granting access to secure facilities or when cross-checking against records during an arrest. Therefore, technical inaccuracies in AI systems or biased outcomes may have negative effects, eventually leading to erroneous or discriminatory outcomes.

In any case, it is noteworthy that, although the AI Act distinguishes between biometric verification and biometric identification systems, both involve the processing of biometric data which enjoy heightened protection pursuant to Article 9 GDPR and Article 10 LED³³⁵. Therefore, while biometric verification systems may be lawfully placed on the market, put into service, and used under the framework of the AI Act, this does not mean that their use is unrestricted. Data protection legislation may still impose additional limitations and safeguards, since biometric verification involves the processing of special categories of personal data, which is generally prohibited unless specific legal grounds and strict conditions are met³³⁶.

³³² European Parliament's Amendments, Amendment 193.

³³³ ESCAJEDO SAN-EPIFANIO, *Biometric Recognition In The Artificial Intelligence Act: Exemptions, Prohibitions And High-Risk Specialties*, cited, 207.

³³⁴ AI Act, Recital 17.

³³⁵ The EDPB has confirmed that authorisation systems relying on biometric data, as defined in Article 4(14) GDPR, fall within the scope of the special protections and restrictions laid down in Article 9 GDPR. This is because biometric authentication, even when carried out on a “one-to-one” basis, generally serves the purpose of identifying the individual. EDPB, *Guidelines 05/022 on the use of facial recognition technology in the area of Law enforcement*, Version 1.0 of 2022 and Version 2.0 of 2023.

³³⁶ Although they both process biometric data, the impact for data protection legislation is different in the two cases. ESCAJEDO SAN-EPIFANIO, *Biometric Recognition In The Artificial Intelligence Act: Exemptions, Prohibitions And High-Risk Specialties*, cited, 204-207.

4. Considerations on the established legal framework

During the legislative process, the EU institutions advanced differing positions, and after extensive negotiations the final text of the AI Act introduced a complex regulatory framework governing the placing on the market, putting into service, and use of AI systems for law enforcement purposes. As analysed above, certain AI practices have been deemed unacceptable within the European Union and consequently subjected to prohibition³³⁷. The decision to ban particular systems or applications is far from straightforward, as it entails an *ex ante* balancing of competing interests of the highest order. Within the operational sector considered in the present Thesis, the imperative of safeguarding public security and public order must be carefully weighed against the protection of individual rights and freedoms of the persons affected. Prohibition, therefore, cannot be the result of a merely precautionary stance, rather it must be grounded in a robust justification, supported by a shared understanding among policymakers, stakeholders, and society at large as to why such measures are necessary.

One of the common grounds on which certain AI practices have been prohibited relates to the potential violations they pose to fundamental rights and freedoms³³⁸. Specific AI systems – for instance, some forms of predictive policing or biometric recognition technologies – carry a heightened risk of undermining principles such as fairness and non-discrimination, individual autonomy, freedom of expression, freedom of assembly and association, as well as privacy and data protection. As I analysed in Chapter 2, these technologies, by their very nature, can lead to systemic biases, disproportionate surveillance, and restrictions on individual liberties, thereby threatening the very democratic values that the European Union seeks to safeguard. Furthermore, it is well established that human rights legislation permits interference with fundamental rights and freedoms only under very strict conditions. Article 52 of the CFREU, for instance, provides that limitations on the rights enshrined therein are admissible only when the essence of the right is preserved, and when the restriction complies with the principles of necessity and proportionality³³⁹. Similarly, the ECHR stipulates that interference with a right guaranteed therein is permissible solely if the measure pursues a legitimate aim and if that aim cannot be achieved by alternative means that would impose fewer restrictions on the right in question. Given that certain AI systems affect the very essence of fundamental rights, the

³³⁷ For a mapping on the debate on prohibited AI practices see C. RUDSCHIES, I. SCHNEIDER, *The Long and Winding Road to Bans for Artificial Intelligence: From Public Pressure and Regulatory Initiatives to the EU AI Act*, in *Digital Society*, 4(57), 2025. E. SELINGER, W. HARTZOG, *The incontestability of facial surveillance*, in *Loyola Law Review*, 66, 2019, 101.

³³⁸ In this terms RUDSCHIES, SCHNEIDER, *The Long and Winding Road to Bans for Artificial Intelligence: From Public Pressure and Regulatory Initiatives to the EU AI Act*, cited. On this specific topic see also A. RENDA ET AL., *Study to support an impact assessment of regulatory requirements for Artificial Intelligence in Europe Final report*, April 2021.

³³⁹ FRA, *Getting the future right - artificial intelligence and fundamental rights*, Publications Office of the European Union, 2020.

European legislature has appropriately deemed them unacceptable and thus prohibited their development and/or deployment within the European legal order³⁴⁰.

A second ground for prohibition arises from broader ethical concerns³⁴¹. Some uses of AI, regardless of their technical feasibility or efficiency, are deemed unacceptable because they conflict with ethical principles. These ethical considerations go beyond strictly legal rights and capture scenarios in which the deployment of AI would be fundamentally incompatible with social norms and moral boundaries. Luciano Floridi argues that an ethical approach to what he calls the “infosphere”³⁴² is necessary because, while compliance with legislation is essential, it is insufficient to guide society in the right direction³⁴³. Legality, therefore, while representing an essential component for the responsible design and development of AI, does not appear to be enough³⁴⁴.

In particular, the term “AI ethics”³⁴⁵ refers to a multidisciplinary field of study where experts debate issues such as the role that AI systems should play in our societies, the risks they entail, how they

³⁴⁰ The attention to fundamental is also reflected in the growing body of literature on digital constitutionalism. This scholarship explores how constitutional values should be reinterpreted and reinforced in the digital age, where the pervasive role of technology challenges traditional legal frameworks. Digital constitutionalism thus provides a normative lens through which to evaluate the development and deployment of AI systems, highlighting the need to embed constitutional guarantees and ethical standards directly into technological governance structures. E. CELESTE, *Digital constitutionalism: The role of internet bills of rights*, Routledge, 2022. G. DE GREGORIO, *Digital Constitutionalism in Europe. Reframing Rights and Powers in the Algorithmic Society*, Cambridge University Press, 2022. For a critical vision of digital constitutionalism see A.J. GOLIA, *Critique of digital constitutionalism: Deconstruction and reconstruction from a societal perspective*, in *Global Constitutionalism*, 13(3), 2024, 488-518. Doi: 10.1017/S2045381723000126.

³⁴¹ In these terms RUDSCHIES, SCHNEIDER, *The Long and Winding Road to Bans for Artificial Intelligence: From Public Pressure and Regulatory Initiatives to the EU AI Act*, cited.

³⁴² L. FLORIDI, *Soft Ethics and the Governance of the Digital* in *Philosophy & Technology*, 31, 2018, 1. With this concept, Floridi seeks to describe the current historical moment in which we no longer live merely online or offline, but *onlife*, that is, we increasingly inhabit that special space where the analog and digital, the offline and online, seamlessly merge. L. FLORIDI, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Raffaello Cortina Editore, 2017. L. FLORIDI, *Pensare l'infosfera. La filosofia come design concettuale*, Raffaello Cortina Editore, 2020.

³⁴³ HLEG, *Ethics Guidelines on Trustworthy AI*, cited, 8. The debate on the ethical implications raised by the use of AI can be traced back to the early 1960s when Norbert Wiener clearly expressed concerns related to the construction of machines capable of autonomously learning and making decisions that humans would not be able to predict and control. Shortly afterwards, Arthur Samuel published an article in which he objected to the fears shown by his contemporaries, stating that machines are mere “*inanimate assemblages of mechanical and electrical parts*”, devoid of any will and, as such, unsuitable to pose a threat to mankind. However, the progress that has been made in the field of AI since 1960 cannot be ignored, and the concerns raised at that time are coming back onto the scene stronger than ever. Therefore, such progress calls for careful reflection on the future we want from a humanity that is increasingly technological and flanked by intelligent systems. See J. MORLEY, L. FLORIDI, L. KINSEY, A. ELHALAL, *From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices*, in *Science and Engineering Ethics*, 26, 2020, 2141. N. WIENER, *Some Moral and Technical Consequences of Automation*, in *Science*, 131 (3410), 1960, 1355-1358. ARTHUR SAMUEL, *Some Moral and Technical Consequences of Automation - A Refutation*, in *Science*, 132 (3429), 1960, 741.

³⁴⁴ In these terms, A. MANTELERO, *AI and big data: A blueprint for a human rights, social and ethical impact assessment*, in *Computer Law & Security Review*, 34(4), 2018, 754-772.

³⁴⁵ According to Floridi, Ethics can essentially take two paths: that of hard ethics and that of soft ethics. Hard ethics concerns values, rights, duties, and responsibilities and, more generally, refers to what is morally right or wrong. Hard ethics is the kind of Ethics that drives the formulation of new regulations or the reconsideration of existing ones. Soft ethics, on the other hand, while operating on the same normative foundation, considers what should or should not be done beyond and in addition to existing regulation. It is, therefore, a post-compliance ethics. FLORIDI, *Soft Ethics and the Governance of the Digital*, cited, 4. In a later work, Floridi sought to identify when it is better to adopt a hard ethics approach and when, instead, a soft ethics approach is preferable. L. FLORIDI, *Soft Ethics: Its Application to the General Data Protection*

should be governed, and which principles should guide research and development in AI³⁴⁶. Addressing the issues raised by AI with an ethical approach could bring several advantages: on the one hand, it could help identify new opportunities that are socially acceptable and desirable; on the other hand, it could help prevent and mitigate socially undesirable or unacceptable practices³⁴⁷. While not all AI ethics guidelines have explicitly called for the prohibition of certain AI practices, many have nonetheless identified specific AI systems as indefensible for development and use, particularly those involving profiling and oppressive forms of surveillance³⁴⁸.

In practice, these two grounds – fundamental rights protection and ethical imperatives – are deeply intertwined. Violations of rights often carry ethical consequences, while ethical judgments provide the normative basis for determining the scope and meaning of rights in a technologically evolving context. The European legislator, in drafting the AI Act, was acutely aware of this interdependence³⁴⁹. The AI

Regulation and Its Dual Advantage in Philosophy & Technology, 31, 2018, 166-167. S. VALLOR, B. GREEN, I. RAICU, *Overview of Ethics in Tech Practice*, Document part of the Ethics in Technology Practice project of the Marrkula Center for Applied Ethics, 22 June 2018, 3.

³⁴⁶ In 2016, the Institute of Electrical and Electronics Engineers launched an initiative to recommend the adoption of policy guidelines aimed at promoting ethically aligned design in AI. Following this initiative, there has been a proliferation of documents containing principles and guidelines, drafted by governments, civil society, and multi-stakeholder groups, all seeking to ensure the proper and responsible development and use of AI. Following this initiative, there has been a proliferation of documents containing principles and guidelines, drafted by governments, civil society, and multi-stakeholder groups, all seeking to ensure the proper and responsible development and use of AI. Within just a few years, dozens upon dozens of principles, codes, and guidelines on AI ethics have been published. On the one hand, these contributions have had positive outcomes: they symbolized a growing awareness of the ethical issues inherent in AI applications and provided important suggestions on how AI should be built and used. On the other hand, they have also created an environment of incoherence and confusion among stakeholders. First, there has been the risk of redundancy, with principles sometimes conflicting with one another and therefore incompatible, rather than offering common, clear, and socially accepted ethical standards to guide both public and private actors. Second, another risk has been the emergence of a true “market of principles and values”, where different stakeholders might simply choose those that best justify their existing behaviour, instead of adopting those that encourage real revision and change in practice. CEPEJ, *Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*, Strasbourg, 3 December 2018, CEPEJ(2018)14. ITU, *AI for Good*. Available at: <https://aiforgood.itu.int/>. ACM, Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 2018 and 2020. Future Of Life Institute, *Asilomar AI Principles*. Available at: <https://futureoflife.org/ai-principles/>. Floridi has identified that, alongside the traditional ethical principles of beneficence, non-maleficence, autonomy, and justice, a further principle specific to AI systems can be added: explicability. L. FLORIDI, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Raffaello Cortina Editore, 2022.

³⁴⁷ This, in turn, would enable companies and organisations to avoid particularly costly risks, especially in terms of public acceptance. In both cases, an ethical approach would help reduce the opportunity costs caused by choices not made or opportunities missed out of fear of making mistakes.

³⁴⁸ From international organisations to civil society groups have called out and condemned certain AI practices as fundamentally unacceptable. UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, 2020. DEK, Opinion of the Data Ethics Commission. Data ethics commission of the federal government of Germany, federal ministry of the interior, building and community, federal ministry of Justice and consumer protection, 2019. Available at: https://www.bfdi.bund.de/SharedDocs/Downloads/EN/Datenschutz/Data-Ethics-Commission_Opinion.pdf?__blob=publicationFile&v=1. University of Montréal, *Montréal declaration for a responsible development of artificial intelligence*, 2018.

³⁴⁹ The development of a secure, trustworthy and ethical AI has been specifically requested by the European Council and the European Parliament. European Council, *Special meeting of the European Council (1 and 2 October 2020) — Conclusions*, EUCO 13/20, 2020, 6. European Parliament, *Resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies*, 2020/2012(INL).

Act is not merely a technical regulatory framework but is grounded in ethical principles that both underpin its legal provisions and guide their interpretation³⁵⁰. Ethics provides conceptual tools to identify and evaluate human actions, yet only some of these can be translated into technical solutions. Others – particularly those closely tied to morality – resist such translation³⁵¹. Nevertheless, the purpose of ethics in relation to technology remains clear: to steer technological development toward virtuous and socially acceptable outcomes³⁵². This involves drawing attention to the kind of people we want to be, the kind of society we want to build, the kinds of technologies we want to develop, how they should be developed and for what purposes³⁵³.

Another basis for prohibition lies in sector-specific risks³⁵⁴. Certain operational domains involve particularly high risks when deploying AI systems, with law enforcement being a prime example. Indeed, many of the AI Act's prohibitions relate directly to this sector. Decisions in such contexts – alongside other sensitive areas including migration and border control, justice, and social welfare – can have especially severe consequences, as they profoundly affect individuals' lives. This is the case with biometric identification systems, where the concerns highlighted in Chapter 2 have led the European legislator to ban such systems when particular conditions are met, such as remote identification, real-time operation, and use in publicly accessible spaces.

Civil society organisations have been among the strongest advocates for prohibiting AI practices considered most threatening to democratic values, such as mass biometric surveillance, predictive policing, and social scoring³⁵⁵. Their calls have been grounded mostly in the above considerations. However, the prohibitions ultimately enshrined in the final text of the AI Act fall short of these demands. While the AI Act does introduce bans on certain practices, these are often limited in scope,

³⁵⁰ This is evidenced by the fact that the AI Act recalls the 2019 *Ethics guidelines for trustworthy AI*. Recitals 7, 27, 165 and Article 95.

³⁵¹ Scientific Foresight Unit (STOA), *Artificial intelligence: From ethics to policy*, 13.

³⁵² Ethics is mainly focused on the identification of principles and rules that govern human behaviour. Ethics is not static; rather, it is an ever-changing discipline. “*Ethics cannot be approached like mathematics; there is no algorithm for ethics, and moral life is not a well-defined, closed problem for which one could design a single, optimal solution. It is an endless task of skilfully navigating a messy, open-ended, constantly shifting social landscape in which we must find ways to maintain and support human flourishing with others, and in which novel circumstances and contexts are always emerging that call upon us to adapt our existing ethical heuristics, or invent new, bespoke ones on the spot*”. S. VALLOR, B. GREEN, I. RAICU, *Technology and Engineering Practice: Ethical Lenses to Look Through*, Document part of the Ethics in Technology Practice project of the Marrkula Center for Applied Ethics, Santa Clara University, 13 July 2020, 8.

³⁵³ The AI Act recognises this in Article 1 paragraph 1: “*The purpose of this Regulation is to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union and supporting innovation*”.

³⁵⁴ In this terms RUDSCHIES, SCHNEIDER, *The Long and Winding Road to Bans for Artificial Intelligence: From Public Pressure and Regulatory Initiatives to the EU AI Act*, cited.

³⁵⁵ Access Now et al., *Open letter calling for a global Ban on biometric recognition technologies that enable Mass and discriminatory surveillance*, 2021. Available at: <https://www.amnesty.org/en/documents/doc10/4254/2021/en/>. Access Now et al., *Joint statement: The EU AI Act must protect people on the move*, 2022. Available at: <https://www.accessnow.org/press-release/joint-statement-ai-act-people-on-the-move/>.

subject to exceptions, or framed in ways that leave significant room for interpretation and circumvention to Member States³⁵⁶.

As discussed above, for example, in the case of person-based predictive policing systems the mere intervention of a human operator, provided that such intervention meets the characteristics previously outlined, is sufficient to remove the system from the scope of prohibition and to reclassify it as a high-risk system. As noted above, the interpretation of the adverb “*solely*” is of central importance. In its judgment *OQ v Land Hessen*³⁵⁷, the CJEU provided important clarification on the notion of “*solely automated processing*”. The Court held that the automated establishment of a probability score concerning an individual’s likelihood of meeting future payment obligations constitutes automated individual decision-making where that score is subsequently relied upon in a decisive manner by third parties, for instance, in assessing the granting of a loan. Crucially, the Court underscored that the concept of “decision” under Article 22 must be interpreted broadly, so as to encompass profiling outputs such as credit scores, given their capacity to produce significant effects for data subjects. This reasoning is particularly instructive beyond the credit sector: by analogy, it may inform the legal assessment of predictive policing tools, where probability scores are generated and transmitted to law enforcement authorities, which then act upon them in ways that may significantly affect individuals. Similar considerations may be emphasised regarding automated video surveillance and, more specifically, biometric identification systems. As discussed above, an absolute prohibition has been transformed through the legislative process into a conditional regime³⁵⁸. Consequently, real-time RBI systems may now be deployed under a series of exceptions, which hardly can be defined “*narrowly defined situations*”³⁵⁹.

These exceptions have been the subject of considerable criticism in the academic debate. Some scholars argue that they open the door for Member States to equip public spaces with camera systems enhanced by biometric technologies, which, once installed, could be reconfigured and activated for the

³⁵⁶ Access Now, *The EU AI act: A failure for human rights, a victory for industry and law enforcement*, 2024. Available at: <https://www.accessnow.org/press-release/ai-act-failure-for-human-rights-victory-for-industry-and-law-enforcement/>.

Amnesty International, *EU: Artificial intelligence rulebook fails to stop proliferation of abusive technologies*, 2024. Available at: <https://www.amnesty.eu/news/eu-artificial-intelligence-rulebook-fails-to-stop-proliferation-of-abusive-technologies/>.

³⁵⁷ CJEU, Judgement of 7 December 2023, *OQ v Land Hessen*, Case C-634/21, ECLI:EU:C:2023:957. L. COTINO HUESO, *La primera sentencia del Tribunal de Justicia de la Unión Europea sobre decisiones automatizadas y sus implicaciones para la protección de datos y el Reglamento de inteligencia artificial*, in *Diario La Ley*, January 2024. Available at: <https://diariolaley.laleynext.es/dll/2024/01/17/la-primera-sentencia-del-tribunal-de-justicia-de-la-union-europea-sobre-decisiones-automatizadas-y-sus-implicaciones-para-la-proteccion-de-datos-y-el-reglamento-de-inteligencia-artificial>.

³⁵⁸ “In view of this, was there really any need for such a disorderly drafting of point (h), which begins rather bluntly only to be progressively emptied of content in its subparagraphs, and then in Article 5.2 to 8?”. LEIRE ESCAJEDO SAN-EPIFANIO, *Biometric Recognition In The Artificial Intelligence Act: Exemptions, Prohibitions And High-Risk Specialties*, cited, 224.

³⁵⁹ AI Act, Recitals 33 and 34.

purposes of RBI whenever the legally defined exceptional circumstances are invoked³⁶⁰. In practice, this raises the risk of “*function creep*”³⁶¹, discussed in Chapter 2, whereby technologies initially introduced under narrow and specific conditions gradually expand into broader, routine applications. Other scholars emphasise that such exceptions may weaken the overall effectiveness of the prohibition itself, thereby undermining the intended safeguard for fundamental rights and freedoms³⁶². Therefore, by allowing conditional use, the regulatory framework risks sending a mixed message: while seeking to prohibit RBI in principle, it simultaneously legitimizes its use under circumstances that could be broadly interpreted. The conditional admissibility of RBI systems thus risks facilitating the gradual normalization of its deployment, creating a slippery slope whereby practices initially tolerated only in exceptional situations become increasingly accepted and entrenched in everyday law enforcement activities.

Despite the failure to introduce a total ban on real-time RBI systems, the analysis above examined the safeguards that the AI Act has introduced for the exceptional cases in which such systems may be deployed³⁶³. In this author’s assessment, whether these measures can be regarded as providing adequate protection for individuals will largely depend on the choices made by Member States, specifically, on whether they permit the use of such systems at all, and if so, under what precise conditions, safeguards, and limitations. Therefore, the way the legal provisions authorising their use will be formulated will ultimately determine whether the balance between interests at stake is effectively maintained or whether the door is left open to disproportionate interference with individual’s fundamental rights and freedoms.

Moreover, Member States are afforded discretion in designating the competent authority responsible for granting authorisations in cases where the use of RBI systems is allowed. They may choose either a judicial authority or an independent administrative authority. Some commentators argue that, while judicial oversight provides stronger safeguards – given that independence is intrinsic to the judiciary

³⁶⁰ N. A. SMUHA, K. YEUNG, *The European Union’s AI Act: beyond motherhood and apple pie?*, cited, 237.

³⁶¹ *Ibidem*.

³⁶² Access Now, *The EU AI act: A failure for human rights, a victory for industry and law enforcement*, cited; European Center for Not-for-profit Law, *Packed with loopholes: Why the AI Act fails to protect civic space and the rule of Law*, 2024. Available at: <https://ecnl.org/news/packed-loopholes-why-ai-act-fails-protect-civic-space-and-rule-law>. L.L. CABRERA, *EU’s much-heralded AI Act agreed by EU Parliament - but serious human rights holes in law remain*, in Euractiv, 2024. Available at: <https://www.euractiv.com/opinion/eus-much-heralded-ai-act-agreed-by-eu-parliament-but-serious-human-rights-holes-in-law-remain/>. Algorithm Watch, *EU’s AI Act fails to set gold standard for human rights*, 2024. Available at: <https://algorithmwatch.org/en/ai-act-fails-to-set-gold-standard-for-human-rights/>; L.L. CABRERA, *EU AI act brief - Pt. 2, privacy & surveillance*, Center for democracy and technology, 30 April 2024. Available: <https://cdt.org/insights/eu-ai-act-brief-pt-2-privacy-surveillance/>; P. HACKER, *Comments on the Final Trilogue Version of the AI Act*, 23 January 2024. Available at SSRN: <https://ssrn.com/abstract=4757603> or <http://dx.doi.org/10.2139/ssrn.4757603>.

³⁶³ RUDSCHIES, SCHNEIDER, *The Long and Winding Road to Bans for Artificial Intelligence: From Public Pressure and Regulatory Initiatives to the EU AI Act*, cited, 57.

– the independence of administrative authorities is less certain³⁶⁴. While this concern does not apply, from my perspective, to independent administrative authorities, it may arise in relation to other authorisation-granting bodies – such as certain market surveillance authorities designated by Member States – which may not meet the higher independence standards the AI Act requires for the law enforcement sector³⁶⁵. This institutional variability thus risks creating uneven levels of protection across the Union, depending on how each Member State structures and safeguards its oversight mechanisms.

A significant question arises with regard to systems that recognise “something”, such as certain characteristics of the individual, within a video feed or an image. An example is an AI system designed to detect suspicious behaviour in a crowd. Since such systems are neither used to identify nor to verify a specific person, they may be understood instead as categorising individuals. In this context, the AI system effectively divides people into those who behave “normally” and those who exhibit suspicious conduct. As some scholars have observed, this interpretation offers stronger protection for individuals, as it ensures that, depending on the categories employed, such AI systems will, in many cases, fall either under the prohibitions set out in Article 5 or, where relevant, within the high-risk classification of Annex III³⁶⁶.

In general, beyond the outright prohibition of certain AI practices, AI providers may face considerable difficulty in accurately categorising the risk level of their systems. This concern is far from theoretical. In practice, significant uncertainties have already emerged surrounding the application of the high-risk classification, particularly given the current absence of detailed guidance from the AI Office³⁶⁷. Without clear and authoritative interpretive benchmarks, providers are left to navigate an inherently complex regulatory framework largely on the basis of their own assessment, raising the prospect of a high degree of subjectivity into the assessments and inconsistent classification decisions across the market. Until the AI Office issues comprehensive guidance, or the criteria are clarified through enforcement practice or judicial interpretation by the CJEU, this ambiguity is likely to persist, potentially undermining both legal certainty for providers and the uniform application of the AI Act across Member States.

Despite the concerns highlighted above, this Regulation is the product of lengthy political negotiations involving concessions between Member States, EU institutions and industry stakeholders. Indeed,

³⁶⁴ In these terms see HACKER, *Comments on the Final Trilogue Version of the AI Act*, cited.

³⁶⁵ AI Act, Article 74(8).

³⁶⁶ ESCAJEDO SAN-EPIFANIO, *Biometric Recognition In The Artificial Intelligence Act: Exemptions, Prohibitions And High-Risk Specialties*, cited, 209.

³⁶⁷ Diverging views and persisting uncertainties surrounding the high-risk classification have been highlighted in a report published by the FRA. FRA, *Assessing High-Risk Artificial Intelligence*, 26.

CHAPTER 3

State authorities often defend surveillance technologies by framing them as necessary and proportionate for public security. The challenge lies thus in deciding what necessary and proportionate truly means, especially when fundamental rights conflict with broader societal interests. While some scholars contend that security benefits rarely outweigh the social and individual harms³⁶⁸, the ultimate judgment on this balance should rest with constitutional courts. This explains why some of the most ambitious proposals, particularly those strongly supported by civil society organisations, for a robust, rights-based framework were either weakened or omitted from the final text, which inevitably reflects a political compromise rather than a principled stance firmly grounded in the protection of fundamental rights³⁶⁹.

³⁶⁸ E. SELINGER, W. HARTZOG, *The incontestability of facial surveillance*, cited. L. STARK, *Facial recognition is the plutonium of AI*, XRDS: Crossroads, The ACM Magazine for Students, 25(3), 2019, 50–55.

³⁶⁹ HACKER, *Comments on the final Trilogue version of the AI Act*, cited, 2; RUDSCHIES, SCHNEIDER, *The Long and Winding Road to Bans for Artificial Intelligence: From Public Pressure and Regulatory Initiatives to the EU AI Act*, cited, 21.

CHAPTER 4

Transparency and fairness in AI-driven law enforcement: a critical analysis under the EU AI Act

1. Overview of the Chapter	271
Part I: Transparency issues in the AI Act	274
1. Transparency in AI systems.....	274
2. The multiple dimensions of transparency in the AI Act	277
2.1. The first dimension of transparency: information obligations	278
2.2. The second dimension of transparency: traceability	282
2.3. The third dimension of transparency: interpretability	283
2.4. The fourth dimension of transparency: explainability.....	287
3. Considerations on the established legal framework with a focus on law enforcement	289
3.1. Technical barriers	297
3.2. Technological barriers	300
3.3. Human oversight	307
3.4. The tactical barriers and the difficulty in ensuring the right to an effective remedy	313
Part II: Algorithmic fairness under the AI Act: implications for law enforcement.....	322
1. Algorithmic fairness	322
2. The legal principle of non-discrimination.....	327
3. The AI Act's provisions on bias and non-discrimination	329
3.1. Discrimination as a potential risk of using AI systems: the central role of a Fundamental Rights Impact Assessment	330
3.2. From data to discrimination: how the AI Act confronts bias in AI-powered law enforcement	339

1. Overview of the Chapter

The law enforcement sector constitutes a particularly sensitive domain within the broader legal framework, where the stakes are high and the equilibrium between the protection of constitutional rights and freedoms, on the one hand, and public security and public order, on the other, is especially fragile. The introduction of AI into this domain risks destabilising this already delicate balance, potentially producing a shift in favour of public security at the expense of individual rights and freedoms¹. It is therefore imperative that the legal system establishes robust safeguards to ensure that

¹ AI Act, Recital 59.

the efficiency – whether real or perceived – derived from the deployment of AI systems is appropriately counterbalanced by mechanisms that uphold fundamental rights and democratic values. The AI Act constitutes the European Union’s principal attempt to provide such a regulatory response. In Chapter 3 I analysed the provisions of the AI Act that govern AI systems placed on the market, put into service, and used for law enforcement purposes. The framework establishes a complex structure that requires law enforcement authorities to carefully assess their roles and responsibilities. Beyond cases involving prohibited AI practices – where AI development and/or deployment are banned – authorities using high-risk AI systems must first determine whether they act as AI system providers. If so, they must comply with the conformity requirements in Chapter III, Section 2, and fulfil the obligations outlined in Articles 16 to 22. Where authorities act as deployers, they are bound by the obligations set forth in Article 26. In general, law enforcement authorities act as deployers of AI systems, given that they typically use technologies that have been developed and supplied by external actors, as discussed in Chapter 2. However, exceptions exist. In some instances, police departments or related agencies may assume the role of providers, particularly when AI systems are developed internally for operational purposes. In such cases, law enforcement authorities assume a dual responsibility, acting simultaneously as deployers and providers of AI systems, and are therefore directly accountable for ensuring compliance also with the obligations set out in Chapter III, Section 3.

Chapter 4 undertakes an in-depth examination of a number of central requirements and obligations that are of particular relevance to the development and deployment of AI systems in the law enforcement context in a manner consistent with fundamental rights and core public values². The Chapter focuses on those regulatory mechanisms that seek to constrain the risks posed by AI technologies while enabling their use within the boundaries of the European legal order. In the first part of the Chapter (Part I), the analysis concentrates on transparency requirements across the AI value chain. It examines, *inter alia*, documentation obligations imposed on providers, information duties owed to deployers and affected persons, and the complex relationship between the legal principle of transparency and technical explainability³. The second part of the Chapter (Part II)

² These concerns are highlighted by all the documents published on the use of AI in law enforcement. Among others, see EUROPOL, *AI and policing The benefits and challenges of AI for law enforcement*, Report from the Europol Innovation Lab, Publications Office of the European Union, 2024. Available at: <https://www.europol.europa.eu/publication-events/main-reports/ai-and-policing>. UNICRI, INTERPOL, *Responsible AI Innovation in law enforcement. AI toolkit*, February 2024. The AI Toolkit is composed by: 00. README File; 01. Introduction to Responsible AI Innovation; 02. Principles for Responsible AI Innovation; 03. Organizational Roadmap; 04. Organizational Readiness Assessment Questionnaire; 05. Risk Assessment Questionnaire; 06. Responsible AI Innovation in Action Workbook; 07. Technical Reference Book. More generally, these issues are prominently addressed in most ethical guidelines and regulatory initiative, and both topics have received considerable attention in the scientific literature, not only within the field of computer science but also in legal scholarship.

³ The importance of transparency issues is shown by the AI Index Report 2024, which indicated that since 2019 the number of papers on transparency and explainability has tripled. Stanford University, *AI Index Report 2024*, 2024, 181.

CHAPTER 4

addresses the principle of fairness, focusing on the ways in which bias – whether embedded in training data or generated and amplified through algorithmic processes – can be identified, mitigated, and ultimately prevented from undermining the effective protection of fundamental rights.

A preliminary observation underpinning the analysis is that concerns relating to opacity and the propagation or amplification of bias in AI systems are frequently framed as consequences of the intrinsic technical limitations of such systems. While these accounts are not without explanatory value, they offer only a partial understanding of the risks at stake. The challenges associated with AI cannot be reduced to purely technical deficiencies; rather, they must be situated within the broader socio-technical and institutional decision-making processes through which AI systems are designed, developed, and deployed. It is this interaction between technological architectures and organisational practices that ultimately gives rise to the potential for rights-infringing outcomes, and which therefore warrants systematic scrutiny within both legal analysis and regulatory design.

Part I: Transparency issues in the AI Act

1. Transparency in AI systems

Transparency has long been a cornerstone of how public authorities shall carry out their functions⁴. In recent years, transparency has attracted growing attention within the AI literature across multiple sectors, primarily in response to the increasing opacity and complexity of contemporary AI systems, whose functioning often remains unintelligible to all but, in certain cases, a limited group of experts⁵. Transparency has also been recognised as a globally endorsed ethical principle, frequently cited in international, supranational and national governance frameworks and ethical guidelines as a foundational standard for the responsible development and deployment of AI⁶. In addition, the relationship between transparency and the deployment of AI systems in public administrations has been the subject of extensive scholarly debate, with transparency widely acknowledged as a fundamental requirement for the legitimate use of such systems by public authorities⁷. Ensuring

⁴ L. ESPOSITO, *Trasparenza e Accountability nelle Amministrazioni Pubbliche: analisi del significato valoriale dei concetti di Trasparenza e Accountability in un'ottica di Open Government*, in www.ildirittoamministrativo.it, 4 February 2021. F. PATRONI GRIFFI, *La Trasparenza della pubblica amministrazione tra accessibilità totale e riservatezza*, in federalismi.it, 8, 2013.

⁵ J. GROCHOWSKI, J. JABŁONOWSKA, F. LAGIOIA, G. SARTOR, *Algorithmic Transparency and Explainability for EU Consumer Protection*, in *Critical Analysis of Law*, 8(1), 2021, 43-63. In an interesting study related to AI ethics, the principle of transparency/explainability/auditability has been identified as the most recurrent principle. See N. KLUGE CORRÊA, C. GALVÃO, J. W. SANTOS, C. DEL PINO, E. PONTES PINTO, C. BARBOSA, D. MASSMANN, R. MAMBRINI, L. GALVÃO, E. TEREM, N. DE OLIVEIRA, *Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance*, in *Patterns*, 4(10), 2023, Article 100857, 1-21. DOI: [10.1016/j.patter.2023.100857](https://doi.org/10.1016/j.patter.2023.100857).

⁶ See High-Level Expert Group on Artificial Intelligence (HLEG), *Ethics guidelines for trustworthy AI*, 2019. Available at: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>; OECD, *OECD principles on artificial intelligence*, OECD Legal Instruments, 2019. Available at: <https://oecd.ai/en/ai-principles>; UNESCO, *Recommendation on the ethics of artificial intelligence*, United Nations Educational, Scientific and Cultural Organization, 2021. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>. Future of Life Institute, *Asilomar AI Principles*, Beneficial AI 2017 conference, 2017. AP4AI *Framework Blueprint issued in the context of the project Accountability Principles for Artificial Intelligence (AP4AI) in the Internal Security Domain*, coordinated by EUROPOL and CENTRIC. Floridi and other authors identified “explainability” as the fifth ethical principle related to AI. See L. FLORIDI, J. COWLS, *A Unified Framework of Five Principles for AI in Society*, in L. FLORIDI (Ed.), *Ethics, Governance, and Policies in Artificial Intelligence*, Springer, 2021, 5-17. L. FLORIDI, J. COWLS, M. BELTRAMETTI, R. CHATILA, P. CHAZERAND, V. DIGNUM, C. LUETGE, R. MADELIN, U. PAGALLO, F. ROSSI, B. SCHAFER, P. VALCKE, AND E. VAYENA, *AI4People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, in *Minds and Machines*, 28(4), 2018, 689–707. For a critical analysis of “explainability” as the fifth ethical principle see J. FIGUEIREDO NOBRE BRITO CORTESE, F. GAGLIARDI COZMAN, M.P. LUCCA-SILVEIRA, A. FIGUEIREDO BECHARA, *Should explainability be a fifth ethical principle in AI ethics?*, in *AI and Ethics*, 3, 2023, 123-134.

⁷ The use of AI systems by public administrations cannot result in making their activities inscrutable. Among the general principles for the use of AI in public administration set out in the “*Piano Triennale per l'informatica nella Pubblica Amministrazione*”, transparency is paramount. Specifically, public administrations must pay particular attention to transparency and interpretability of AI models to ensure accountability, and they must report on decisions made with the support of AI technologies. The importance of transparency when AI systems are used by public administrations has been widely acknowledged in the administrative literature. See L. TORCHIA, *Lo Stato digitale. Una introduzione*, il Mulino, 2025. B. MARCHETTI, *Amministrazione digitale*, in *Enciclopedia del Diritto: funzioni amministrative*, Giuffrè, Milano, 2022, 75-109; M.B. ARMIENTO, *Pubbliche amministrazioni e intelligenza artificiale. Strumenti, principi e garanzie*, Editoriale Scientifica, 2024; G. AVANZINI, *Decisioni amministrative e algoritmi informatici*, Editoriale Scientifica, 2018.

transparency entails safeguarding the accountability of public authorities, reinforcing public trust⁸ and enabling individuals to exercise their fundamental procedural rights, such as the right to challenge the outcome of the AI system and request remedies.

While ensuring transparency in AI systems yields evident benefits, its realization remains a highly complex and multifaceted challenge. The difficulty arises primarily from the intrinsic opacity of AI-based software, an opacity that, more broadly, characterizes software in general. This opacity manifests on several levels. First, it is linguistic: as discussed in Chapter 1, algorithms are encoded in formal programming languages that are inaccessible to those without technical expertise, thereby rendering their operation opaque to non-specialists⁹. Second, it is legal: software is often shielded by intellectual property rights and trade secret protections, which restrict public disclosure and scrutiny¹⁰. Third, it is structural: the increasing sophistication and complexity of contemporary AI models make their inner workings difficult to interpret even for experts. As will be discussed further in this Chapter, ensuring transparency in the context of law enforcement is even more complex, as it must be balanced against compelling interests, such as safeguarding public security, preserving the effectiveness of investigations, and protecting sensitive data.

In addition, considerable ambiguity persists regarding the very object of transparency, *i.e.* whether it concerns the training data, the AI model, the entire decision-making process or the outcomes of AI systems. Moreover, ambiguity persists regarding the allocation of responsibility for ensuring transparency, that is which actors are tasked with its provision, and the intended audience or stakeholders to whom transparency should be directed. Furthermore, there is a lack of clarity regarding the specific objectives that transparency is intended to serve, such as enabling accountability, fostering public trust, facilitating error detection, or safeguarding procedural rights¹¹. Equally uncertain are the mechanisms and instruments through which transparency might be effectively operationalised¹², as well as the normative and technical benchmarks against which its adequacy should be evaluated. Most importantly, a review of policy documents and academic literature reveals that no universally accepted terminology exists to designate efforts aimed at

⁸ By contrast, the lack of transparency generally creates distrust T.M. PETERS, R.W. VISSER, *The importance of distrust in AI*, in *Explainable Artificial Intelligence First World Conference, xAI 2023, Lisbon, Portugal, July 26-28, 2023, Proceedings, Part III (CCIS, volume 1903)*, arXiv:2307.13601, 2023.

⁹ This is well explained by G. SARTOR, *L'informatica giuridica e le tecnologie dell'informazione. Corso d'informatica giuridica*, Giappichelli, 2022.

¹⁰ For a discussion on the protection of software see A. BLANDINI, *Diritto dell'innovazione*, Wolters Kluwer, CEDAM, 2022.

¹¹ Transparency is widely recognised as a *foundational prerequisite* for responsible AI innovation, underpinning key aspects such as accountability, human control and oversight, and the ability to assess the robustness, safety, and accuracy of AI systems. Cfr. UNICRI, INTERPOL, *AI Toolkit – 0.2. Principles for Responsible AI Innovation*, cited, 19.

¹² The National Institute of Standards and Technology (NIST) identified four principles which represent the fundamental properties for explainable AI systems, namely explanation, meaningful explanation, explanation accuracy and knowledge limits. Cfr. NIST, *Four Principles of Explainable Artificial Intelligence*, NISTIR 8312, September 2021. Available at: <https://doi.org/10.6028/NIST.IR.8312>.

rendering AI systems “transparent”¹³. Terms such as “transparency”¹⁴, “interpretability”¹⁵, “explainability”¹⁶ or “explicability”¹⁷ or various combinations thereof¹⁸ are frequently employed in both policy and academic discourse to give concrete expression to the legal and ethical principle of transparency¹⁹.

For the purposes of the present Thesis, transparency will be understood in a broad sense: transparency is conceived as a governance-oriented principle that entails disclosing information about the AI system, both in isolation and as an integral element of the wider socio-technical decision-making process²⁰. Such a broad conception of transparency is cogently articulated by Balint Gyevnara, Nick

¹³ The enduring ambiguity of the transparency principle is evident in D. SCHNEEBERGER ET AL., *The Tower of Babel in Explainable Artificial Intelligence (XAI)*, in A. HOLZINGER, P. KIESEBERG, F. CABITZA, A. CAMPAGNER, A.M. TJOA, E. WEIPPL (Eds.), *Machine Learning and Knowledge Extraction*, CD-MAKE 2023, Lecture Notes in Computer Science, Volume 14065 Springer, Cham, 2023, 73-81. The definition of transparency enshrined in the *Ethics Guidelines for Trustworthy AI* and recalled in Recital 27 of the AI Act does not precisely define it, rather it identifies its constituent elements: “Transparency means that AI systems are developed and used in a way that allows appropriate traceability and explainability, while making humans aware that they communicate or interact with an AI system, as well as duly informing deployers of the capabilities and limitations of that AI system and affected persons about their rights”.

¹⁴ HLEG, *Ethics Guidelines for Trustworthy AI*, cited; European Commission, *White Paper on Artificial Intelligence: a European approach to excellence and trust*, 2020; IEEE *Ethically Aligned Design, First Edition*, IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2019; Council of Europe, *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, 2024.

¹⁵ Cfr. Z.C. LIPTON, *The mythos of model interpretability*, in *Communications of the ACM*, 61(10), 2018, 36-43. Interestingly, the paper distinguishes between *transparency* (how a model works internally) and *post-hoc explanations* (explaining a model’s outputs after the fact) as “competing notions” of interpretability. DOSHI-VELEZ, B. KIM, *Towards a rigorous science of interpretable machine learning*, 2017, arXiv:1702.08608. Available at: <https://doi.org/10.48550/arXiv.1702.08608>. J.S. CARDOSO, *Machine Learning Interpretability: A Survey on Methods and Metrics*, in *Electronics*, 2019, 8(8), 832. Available at: <https://www.mdpi.com/2079-9292/8/8/832>.

¹⁶ For a comprehensive study on the landscape of explainable AI cfr. A. BARREDO ARRIETA ET AL., *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI*, in *Information Fusion*, 58, 2020, 82-115. The paper positions explainability as crucial for “Responsible AI”, arguing that effective XAI enables fairness and accountability in real-world AI deployments. A. ADADI, M. BERRADA, *Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)*, IEEE Access, 6, 2018, 52141-52142. DOI: 10.1109/ACCESS.2018.2870052.

¹⁷ FLORIDI ET AL., *AI4People - An Ethical Framework for a Good AI Society*, cited.

¹⁸ In some cases, “transparency” and “explainability”. Cfr. OECD, *Recommendation of the Council on Artificial Intelligence*, May 2019. Available at: <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>; UNESCO *Recommendation on the Ethics of Artificial Intelligence*, November 2021. In other cases, “interpretability” and “explainability”. C. RUDIN, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, in *Nature Machine Intelligence*, 2019, 1, 206–215. C. MOLNAR, *Interpretable machine learning: A guide for making black box models explainable*, 2nd ed, 2020. D.V., CARVALHO, E.M. PEREIRA, J.S. CARDOSO, *Machine Learning Interpretability: A Survey on Methods and Metrics*, in *Electronics*, 8(8), 2019, Article 832, 1-34.

¹⁹ Different domains (technical, ethical and legal) tend to use different terminology, and, even within a single domain, a consistent taxonomy is lacking. M.E. GUTIÉRREZ DAVID, *Transparency and provision of information to deployers in Article 13*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 668.

²⁰ This definition is taken from S. BUIJSMAN, *Transparency for AI systems: a value-based approach*, in *Ethics and Information Technology*, 26, 2024, Article 34, 1-15. <https://doi.org/10.1007/s10676-024-09770-w>. Such definition is in line with the definition enshrined in the AI Act “transparency means that AI systems are developed and used in a way that allows appropriate traceability and explainability, while making humans aware that they communicate or interact with an AI system, as well as duly informing deployers of the capabilities and limitations of that AI system and affected persons about their rights”. AI Act, Recital 27. The European Data Protection Supervisor understands “transparency” as “the ability for a specific model to be understood. In the strictest sense, a model is transparent if a person can contemplate the entire model at once” and, further, states that “Transparency can be considered at the level of the entire model, at the level of individual components (e.g., parameters), and at the level of a particular training algorithm. A second and less strict notion of transparency might be that each part of the model (e.g., each input, parameter, and computation) admits

Ferguson, and Burkhard Schafer, who observe that “the law treats transparency as a quality of complex socio-technical interactions between the AI and its users, developers, owners, and wider society”²¹ and further that “[t]ransparency in the view of the law is not a goal in itself, but a means that is needed to promote a range of very different values”²².

In this author’s assessment, this broad definition of transparency seems to adequately address the two main objectives of transparency in AI systems. On the one hand, it enables professional users (*i.e.*, deployers) to make informed decisions regarding both whether to engage with the AI system and how to use it appropriately. On the other hand, transparency guarantees that individuals affected by an AI system are informed of its operation and the role it had within the decision-making process, thereby equipping them to exercise their rights, including the ability to contest outcomes that directly affect them.

2. The multiple dimensions of transparency in the AI Act

Consistent with the objectives discussed above, the AI Act advances transparency across multiple dimensions – information, traceability, explainability, and interpretability²³ – by embedding corresponding obligations and requirements throughout its regulatory framework²⁴. In the following

an intuitive explanation”. Moreover, “A model is considered to be transparent if by itself it is understandable. Since a model can feature different degrees of understandability, transparent models in Section 3 are divided into three categories: simulatable models, decomposable models and algorithmically transparent models”. BARREDO ARRIETA, *Explainable Artificial Intelligence (XAI)*, cited, 85.

²¹ B. GYEVNAR, N. FERGUSON, B. SCHAFER, *Bridging the Transparency Gap: What Can Explainable AI Learn from the AI Act?*, in *Frontiers in Artificial Intelligence and Applications*, 372, 2023, 964. By contrast, technical literature – and XAI in particular – “views transparency merely as an algorithmic property that offers practical solutions but through a limited, technology-focused scope”. The authors state that this different understanding of the same concept give rise to a “transparency gap”. “The transparency gap is also one of the reasons why many of the legal requirements are being criticised by writers from the computer science community as being ineffective, overreaching, or technically infeasible. The mistake here is to assume that the broad and ambitious transparency requirements that the Act lays out are engineering instructions, and can be addressed solely through appropriate design decisions. A much better understanding however is to ask which design decisions facilitate, and which ones hinder, the type of transparency aspiration that the Act aims for, without equating algorithmic explainability with legal transparency outright”.

²² *Ibidem*.

²³ Some authors talk about «multiple layers of transparency». M.E. KAMINSKI, *The Right to Explanation, Explained*, in *Berkeley Technology Law Journal*, 34, 2019, 189. For an interesting discussion on higher transparency of AI systems see M. MARTINI, *Regulating Artificial Intelligence - How to De-Mystify the Alchemy of Code?*, in M. EBERS S. NAVAS (Eds.) *Algorithms and Law*, Cambridge University Press, 2020, 100-135.

²⁴ In the HLEG, *Ethics Guidelines for Trustworthy AI*, cited, 14 three key dimensions of transparency are identified: traceability, explainability, and communication. Interpretability is regarded as a component of explainability, as indicated by the questions included in the *Trustworthy AI Assessment List*: “4. Transparency [...] Explainability. Did you assess: to what extent the decisions and hence the outcome made by the AI system can be understood? to what degree the system’s decision influences the organisation’s decision-making processes? why this particular system was deployed in this specific area? what the system’s business model is (for example, how does it create value for the organisation)? Did you ensure an explanation as to why the system took a certain choice resulting in a certain outcome that all users can understand? Did you design the AI system with interpretability in mind from the start? Did you research and try to use the simplest and most interpretable model possible for the application in question? Did you assess whether you can analyse your training and testing data? Can you change and update this over time? Did you assess whether you can examine

sections, I will examine the four dimensions of transparency in greater detail, referring to the specific provisions of the AI Act that address each of them. In the subsequent section, I will critically evaluate this regulatory approach, with particular attention to its implications for law enforcement.

2.1. The first dimension of transparency: information obligations

First and foremost, information obligations may be understood from two complementary perspectives. The first concerns individual-oriented transparency, which focuses on ensuring that end users²⁵ or affected individuals are adequately informed about their interaction with, or exposure to, AI systems. The second, while related, is conceptually distinct and pertains to institutional transparency, which is directed toward informing the wider public about the existence and deployment of AI systems at a systemic level²⁶.

As far as individual-oriented transparency is concerned, Article 26(11), for high-risk AI systems, and Article 50, for limited-risk AI systems, establish information duties incumbent upon both providers²⁷ and deployers²⁸. In particular, the AI Act mandates that individuals must be informed when they are interacting with an AI system²⁹, ensuring awareness of the AI-based nature of the interlocutor (e.g., chatbots)³⁰, or are exposed to AI systems (e.g., AI-generated content), as well as when an AI system is used to make, or to assist in making, decisions that affect them. Certain exceptions and specific provisions apply in the domain of law enforcement. For high-risk AI systems used for law

interpretability after the model's training and development, or whether you have access to the internal workflow of the model?". See also GUTIÉRREZ DAVID, *Transparency and provision of information to deployers in Article 13*, cited, 684. The author identifies three dimensions. However, interpretability and explainability are treated as distinct concepts throughout the chapter, effectively resulting in four dimensions. The distinction between explainability and interpretability is explicitly emphasized in K.R. VARSHNEY, *Trustworthy Machine Learning*, Independently Published, 2022, 163. The Council of Europe's *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* sets out a series of high-level *Principles related to activities within the lifecycle of artificial intelligence systems* (Chapter III), intended to guide the legislative activity of Member States. Article 8 specifically addresses transparency and oversight, stipulating that: "*Each Party shall adopt or maintain measures to ensure that adequate transparency and oversight requirements tailored to the specific contexts and risks are in place in respect of activities within the lifecycle of artificial intelligence systems, including with regard to the identification of content generated by artificial intelligence systems*".

²⁵ In this context, the concept of "user" should not be conflated with that of the "deployer". The term "user" in this Article refers to the end user, for example, an individual interacting with an AI system such as ChatGPT.

²⁶ A clear example of the implementation of this form of *institutional transparency* can be found in Article 30(5) of the Italian Public Procurement Code, which provides that public administrations shall publish, on their institutional website, in the section "Transparent Administration", the list of technological solutions – including AI systems – that are used for the performance of their activities.

²⁷ AI Act, Article 50, paragraphs 1 and 2.

²⁸ AI Act, Article 50, paragraphs 3 and 4.

²⁹ Carlo Casonato proposes adapting the right to informed consent to the context of artificial intelligence, highlighting the need to reinterpret traditional human rights in light of emerging technologies. C. CASONATO, *Unlocking the Synergy: Artificial Intelligence and (Old and New) Human Rights*, in *BioLaw Journal*, 3, 2023, 235–236.

³⁰ This is considered to be closely linked to the right to human dignity. See European Commission, *Annexes to the Impact Assessment of the Regulation on Artificial Intelligence*, cited, 47. The right of individuals to know when they are interacting with a machine has also been emphasised by Stuart Russell in his editorial "*Rise of AI Debaters Highlights Need for Transparency*", in *Nature*, 592, 2021, 166.

CHAPTER 4

enforcement purposes, the disclosure requirement is governed by Article 13 of Directive (EU) 2016/680³¹, which provides a framework for balancing the right of individuals to be informed with the need to preserve the effectiveness of law enforcement activities, protect public security, and avoid jeopardising ongoing investigations. Accordingly, while the general rule under the AI Act requires transparency towards individuals affected by Annex III systems, in the law enforcement context, the right to information is modulated by the specific safeguards, exceptions, and proportionality considerations enshrined in the LED³².

In relation to institutional transparency, providers are required to register high-risk AI systems enumerated in Annex III – and those that are not considered as high-risk pursuant to Article 6(3) – in the European database established by Article 71 of the AI Act, set up and maintained by the European Commission. The database is specifically intended to ensure transparency by registering the extensive set of information listed in Annex VIII and referred to in the figure below.

³¹ AI Act, Article 26(11).

³² Such a restriction is typically justified by the necessity to preserve the integrity of ongoing investigations, ensure public security, and avoid the risk of compromising evidence or investigative techniques.

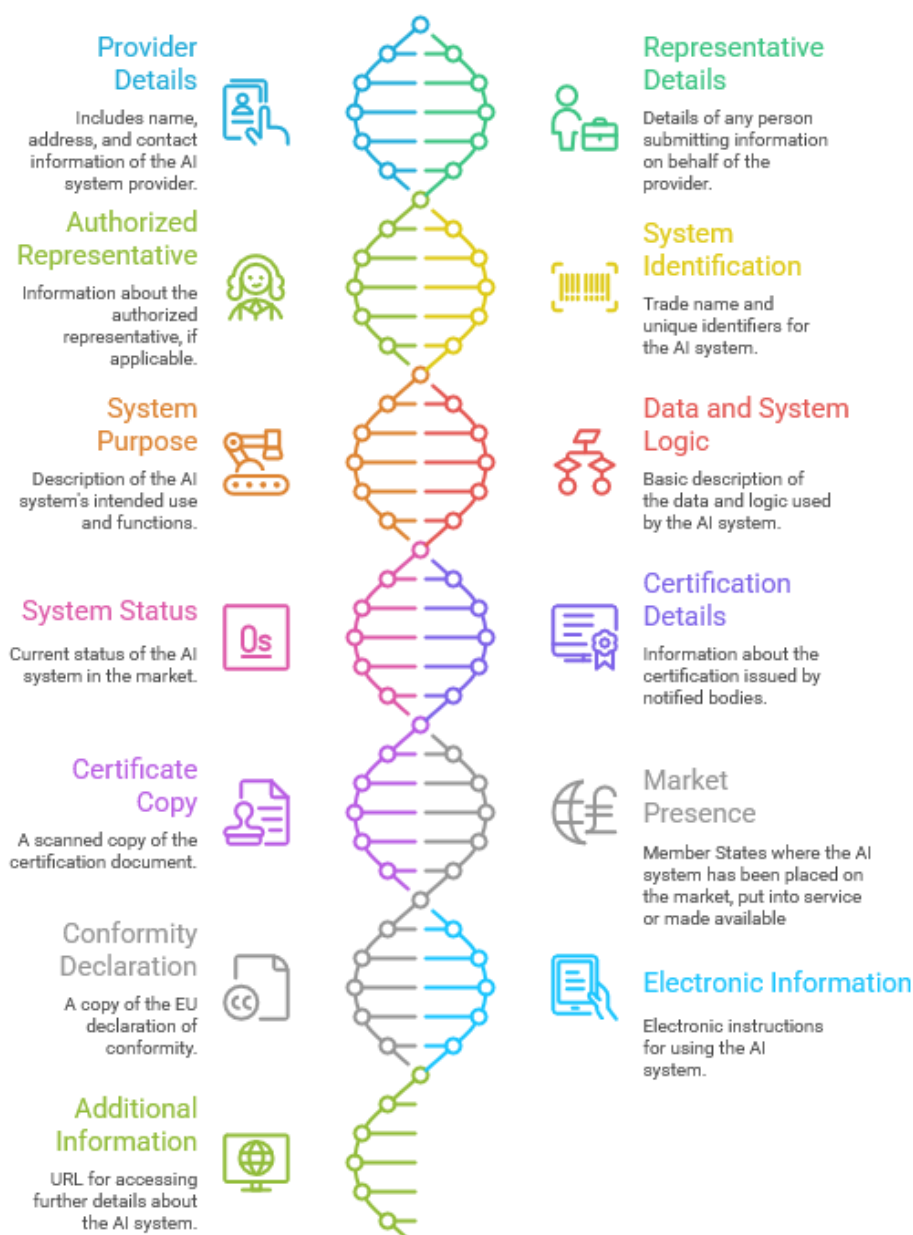


Figure 11: Annex VIII, Section A – Information to be submitted by providers of high-risk AI systems in accordance with Article 49(1). Own elaboration with Napkin.

Similarly, deployers of high-risk AI systems listed in Annex III³³ who are – or who acts on behalf of – a public authority, agency or body, shall register their use of a high-risk AI system, in accordance with Article 49(3). In particular, deployers shall register the name, address and contact details of the deployer or of the person submitting information on behalf of the deployer; the uniform resource locator (URL) of the entry of the AI system in the EU database by its provider; a summary of the findings of the Fundamental Rights Impact Assessment conducted (hereinafter “FRIA”) in

³³ With the exception of high-risk AI systems listed in point 2 of Annex III.

accordance with Article 27 and a summary of the Data Protection Impact Assessment (hereinafter “DPIA”) carried out in accordance with Article 35 of the GDPR³⁴ or Article 27 of the LED³⁵, where applicable.

The information recorded in this database shall, as a rule, be publicly accessible³⁶. By way of derogation, AI systems deployed for the purpose of law enforcement shall be registered in a secure, non-public section of the database, in accordance with Articles 49(4), where only limited information is to be made available. With regard to providers of high-risk AI systems for law enforcement purposes, Article 49(4) specifies that only the information listed in points 1 to 10 of Annex VIII – excluding points 6, 8, and 9 – must be included in the registration³⁷. As a result, using person-based predictive policing systems as an example, a basic and concise description of the data used by the system (such as input data) and its operating logic does not need to be disclosed. Likewise, details such as the type, number, and expiry date of the certificate issued by the notified body, the name or identification number of that body, and, where applicable, the scanned copy of the certificate referred to in point 8 are not required. With regard to deployers, Article 49(4) stipulates that the registration shall be limited to the name, address and contact details of the deployer, or of the person submitting the information on behalf of the deployer and the URL of the entry of the AI system in the database, as created by its provider. By contrast, law enforcement authorities acting as deployers are exempt, unlike other deployers, from the obligation to register a summary of the findings of the FRIA conducted pursuant to Article 27, as well as of the DPIA carried out in accordance with Article 27 of the LED. For the same reason, high-risk AI systems referred to in point 1 of Annex III, which concerns biometric identification and categorisation systems, as well as emotion recognition systems, when

³⁴ For a comment on Article 35 GDPR see A. MANTELERO, *Comment to Article 35 and 36*, in M. COLE, F. BOEHM (Eds.), *GDPR Commentary*, Edward Elgar Publishing, 2019.

³⁵ For a comment of Article 27 of the LED see M. HANSEN, *Article 27 Data protection impact assessment*, in E. KOSTA, F. BOEHM (Eds.), *The EU Law Enforcement Directive (LED): A Commentary*, 2024; Oxford Law Pro, 417-440.

³⁶ The EU database of high-risk AI systems cites Amsterdam and Helsinki, which have launched open AI registers to track how algorithms are used within their municipalities. AI for Good, Helsinki and Amsterdam launch AI registers to detail city systems. Available at: <https://aiforgood.itu.int/helsinki-and-amsterdam-launch-ai-registers-to-detail-city-systems/>. Information about Public AI Registers are available at: <https://openresearch.amsterdam/en/page/73074/public-ai-registers>. I discussed this topic in M. MERLER, *La città nella transizione digitale. AI Localism per una governance dell'AI a misura urbana*, in S. FRANCA, A. PORCARI, S. SULMICELLI (Eds.), *Le transizioni e il diritto*, Quaderni della Facoltà di Giurisprudenza, 2024. A recommendation that the use of AI systems in the public sector, along with relevant information, should be publicly registered is also made by the House of Lords, Justice and Home Affairs Committee, *Technology rules? The advent of new technologies in the justice system*, HL Paper 180, London, 30 March 2022, 44-46. Available at: <https://publications.parliament.uk/pa/ld5802/ldselect/ldjusthom/180/180.pdf>.

³⁷ The following information shall be recorded in a secure, non-public section of the database: 1. The name, address and contact details of the provider; 2. Where submission of information is carried out by another person on behalf of the provider, the name, address and contact details of that person; 3. The name, address and contact details of the authorised representative, where applicable; 4. The AI system trade name and any additional unambiguous reference allowing the identification and traceability of the AI system; 5. A description of the intended purpose of the AI system and of the components and functions supported through this AI system; 7. The status of the AI system (on the market, or in service; no longer placed on the market/in service, recalled); 10. Any Member States in which the AI system has been placed on the market, put into service or made available in the Union.

they are used in law enforcement, migration, asylum, and border control management³⁸, they must be registered in a secure non-public section of the EU database.

Information obligations most closely embody the traditional notion of transparency, which is primarily concerned with the relationship between the one taking the decision – in this case the deployer of AI systems – and the affected individual³⁹. This contrasts with other forms of transparency, such as technical transparency, which is aimed at regulatory authorities through traceability requirements (Articles 11 and 12), and operational transparency, which is aimed at those deploying the AI system so they can understand how it functions and what its capabilities and limitations are (Article 13).

2.2. The second dimension of transparency: traceability

Traceability obligations are enshrined in Articles 11 and 12 of the AI Act, specifically requiring that AI systems are developed and documented in a way that allows their processes, data, and decision pathways to be reconstructed and audited⁴⁰. The AI Act pursues traceability chiefly through obligations concerning the elaboration of technical documentation and systematic record-keeping. These obligations, being primarily technical in nature, are binding mainly on the providers of high-risk AI systems. Moreover, unlike other transparency requirements, the addressees of traceability obligations – namely technical documentation and log record-keeping – are national authorities, rather than deployers or affected individuals⁴¹.

³⁸ Biometric recognition AI systems may likewise be implemented in a wide range of sectors, such as employment, education, healthcare, and the provision of essential services.

³⁹ In these terms G. OLIVATO, *The AI Act: regulatory and technical challenges*, PhD Thesis, University of Trento, 20 June 2025, 262. In similar terms, Ada Lovelace Institute, AI Now Institute and Open Government Partnership, *Algorithmic accountability for the public sector*, 2021, 18. Available at: <https://www.opengovpartnership.org/wp-content/uploads/2021/08/algorithmic-accountability-public-sector.pdf>. In the report it is specified that “Transparency mechanisms provide information about algorithmic systems to the general public (e.g. affected persons, media or civil society) so that individuals or groups can learn that these systems are in use, and demand answers and justifications related to such use”.

⁴⁰ ETSI, *Securing Artificial Intelligence (SAI). Traceability of AI Models*, ETSI TR 104 032 V1.1.1 (2024-02). Available at: https://www.etsi.org/deliver/etsi_tr/104000_104099/104032/01.01.01_60/tr_104032v010101p.pdf. On this topic see F. RAMÓN FERNÁNDEZ, *Quality Management Systems, Technical Documentation And Documentation Keeping In The Regulation*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, 629-647.

⁴¹ As of now, Member States are still in the process of identifying the competent authorities to perform these functions. In Spain, the Spanish Agency for the Supervision of Artificial Intelligence has been appointed as the market surveillance authority, while in Italy, this role has been assigned to the National Cybersecurity Agency. However, the designation of these authorities must be interpreted in light of Article 74, paragraphs 3, 6, and 8 of the AI Act. Specifically, Article 74(8) provides that, for high-risk AI systems listed in point 1 of Annex III – insofar as such systems are used for law enforcement, border management, justice, or democratic processes – and for those listed in points 6, 7, and 8 of the same Annex, Member States shall designate as market surveillance authorities either the competent data protection supervisory authorities under Regulation (EU) 2016/679 (GDPR) or Directive (EU) 2016/680 (Law Enforcement Directive), or any other authority designated in accordance with the conditions laid down in Articles 41–44 of Directive (EU) 2016/680. Given the particular sensitivity of this sector, it is therefore essential to ensure a heightened degree of institutional independence.

Specifically, Article 11 imposes on providers the duty to compile and maintain comprehensive technical documentation for high-risk AI systems. Such documentation must demonstrate conformity with the requirements laid down in Section 2 prior to the system's placement on the market or its entry into service and must be continuously updated. It is required to encompass, inter alia, the general characteristics, capabilities, and limitations of the system, as well as information regarding its algorithms, data, and the processes of training, testing, and validation, in addition to documentation of the applicable risk-management system⁴².

Complementing this, Article 12 establishes log-keeping obligations, requiring providers to ensure that high-risk AI systems are designed in such a way as to enable the automatic recording of events throughout their operational lifetime⁴³. In line with the requirement set out in Article 12, Article 26(6) imposes a corresponding duty on deployers of high-risk AI systems to retain the logs that are automatically generated by such AI systems ("record-keeping"⁴⁴), provided that these logs are within their control. This obligation thus mirrors the provider's duty to ensure that high-risk AI systems are designed and developed with automatic logging capabilities, thereby enabling effective traceability, accountability, and post-market monitoring. By requiring deployers to keep these logs, the Regulation ensures that evidence of the system's functioning is preserved, facilitating both internal compliance checks and external supervisory oversight and supports the broader risk management framework established for high-risk AI systems under the AI Act.

2.3. The third dimension of transparency: interpretability

Interpretability refers to the degree to which a human can understand the internal working of a model and how inputs relate to outputs⁴⁵. In other words, it reflects how comprehensible and intelligible the model is to a human operator⁴⁶. To achieve interpretability two principal approaches can generally be

⁴² The technical documentation shall contain, at a minimum, the elements set out in Annex IV. Where a high-risk AI system related to a product covered by the Union harmonisation legislation listed in Section A of Annex I is placed on the market or put into service, a single set of technical documentation shall be drawn up containing all the information set out in Article 11(1), as well as the information required under those legal acts.

⁴³ For a discussion on this requirement see W. ARELLANO TOLEDO, A. MERCHÁN MURILLO, *The Obligation to Keep Records Of High-Risk Systems In The Artificial Intelligence Act*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, 649-663.

⁴⁴ AI Act, Recital 91.

⁴⁵ Similarly, European Data Protection Supervisor (EDPS), *TechDispatch on Explainable Artificial Intelligence*, #2/2023. Available at: <https://op.europa.eu/s/Aaly>. In some cases, interpretability is confused with explainability, while in others, explainability is regarded as a component of interpretability. GUTIÉRREZ DAVID, *Transparency and provision of information to deployers in Article 13*, cited, 685-686.

⁴⁶ The literature on the topic is huge. Among others, A. SANTOSUOSSO, G. SARTOR, *Decidere con l'IA. Intelligenze artificiali e naturali nel diritto*, Il Mulino, 2024. Z.C. LIPTON, *The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery*, in *ACM Queue*, 16(3), 2018, 12; B. LEPRI ET AL., *Fair, Transparent, and Accountable Algorithmic Decision-Making Processes*, in *Philosophy & Technology*, 31(4), 2018, 619; A. BARREDO ARRIETA ET AL., *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and*

distinguished⁴⁷: one that emphasises the intrinsic design of the system, and another that focuses on the communication of its processes and outcomes to recipients after the results have been generated⁴⁸. More precisely, the first relies on the use of intrinsically transparent models (so called “self-interpretable models”), whose structure is sufficiently interpretable to allow experts to directly analyse and reason about their functioning⁴⁹. In general, simple models based on an *if-then* structure – based on causal relationships, where the software reaches a predefined conclusion once all specified conditions are satisfied – are considered interpretable models, as the reasoning process from input data to output remains intelligible to human operators⁵⁰.

The second approach employs post-hoc, complementary explanations for black-box models in which the internal mechanisms are not human-readable and must therefore be approximated or reconstructed through additional explanatory techniques⁵¹. Examples of black-box models include deep neural networks, which present highly complex structures and rely on statistical correlations, rather than on causal links. Accordingly, explanations of the underlying process are required to allow for a critical assessment of the outputs generated. In conclusion, while some models are inherently interpretable, others lack direct interpretability but can be partially explained through various techniques⁵²,

Challenges toward Responsible AI, 2019, arXiv:1910.10045, 88-100; Information Commissioner’s Office, Alan Turing Institute, *Explaining decisions made with AI*, 2022, 67. Available at: <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/>. B. MITTELSTADT, C. RUSSELL, S. WACHTER, *Explaining Explanations in AI*, in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, 2019, 279-288.

⁴⁷ The literature further emphasises that interpretable models should be understood within the specific domain in which their interpretability is situated. This interpretability is commonly characterised along three key dimensions: algorithmic transparency, decomposability, and simulatability. A. BARREDO ARRIETA ET AL., *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges*, cited. “Algorithmic transparency” refers to the ability of human operators to understand how a model transforms input data into output. In other words, it concerns the user’s capacity to grasp the internal process by which the model produces its results. For instance, “a linear model is deemed transparent because its error surface can be understood and reasoned about, allowing the user to understand how the model will act in every situation it may face. Contrarily, it is not possible to understand it in deep architectures as the loss landscape might be opaque since it cannot be fully observed and the solution has to be approximated through heuristic optimization (e.g. through stochastic gradient descent)”. *Ibidem*, 88. “Decomposability” concerns the interpretability of a model’s individual components, namely its inputs, parameters, and computations. It represents the ability to explain each of these parts in an intelligible way. “Simulatability” captures the extent to which a human can mentally follow and reproduce the model’s reasoning process. It thus depends heavily on the model’s complexity: an interpretable model in this sense is one that can be easily represented and understood through text or visualisations. Even among models generally classified as interpretable (for example, rule-based models), the degree of interpretability can vary substantially. Some models offer full transparency across all three dimensions, while others provide only partial or limited insight into their decision-making processes. *Ibidem*, 87-88.

⁴⁸ See ICO, Alan Turing Institute, *Explaining decisions made with AI*, cited, 184-185. See EDPS, *TechDispatch on Explainable Artificial Intelligence*, cited, 6-9.

⁴⁹ “In “white box” models, the algorithms used are straightforward to understand and it is possible to interpret how the input features are transformed into the output or target variable”. EDPS, *TechDispatch on Explainable Artificial Intelligence*, cited, 6. Interpretability can be provided at different levels, namely for the whole model, for individual components (e.g., parameters), or for the training algorithm.

⁵⁰ Even with interpretable models, specialised expertise is still necessary to correctly understand and assess their outputs.

⁵¹ EDPS, *TechDispatch on Explainable Artificial Intelligence*, cited, 7-8.

⁵² MITTELSTADT, RUSSELL, WACHTER, *Explaining Explanations in AI*, cited, 280.

providing different forms of explanation depending on the intended purpose of the explanation and its audience⁵³.

The AI Act seeks to address interpretability concerns through Article 13(1)⁵⁴. This provision establishes a “*transparency requirement*” for high-risk AI systems, placing primary responsibility on providers, who are mandated to design and develop AI systems in a way that ensures sufficient transparency. In addition, providers are requested to supply the deployer with clear and accessible instructions for use⁵⁵, which must include, among other elements, information enabling deployers to interpret the system’s outputs and to use them appropriately⁵⁶, as well as details on human oversight measures⁵⁷, including the technical mechanisms designed to facilitate the interpretation of the system’s outputs by the deployer.

Designing AI systems to be interpretable is a typical example of the by design approach discussed in Chapter 3. This approach requires that AI systems be designed and developed in a manner that ensures their operation is “*sufficiently transparent*” to allow those responsible for deployment to interpret the output, favouring the development of inherently interpretable models over the continued reliance on opaque systems supplemented by post-hoc explanations. Through mandating transparency mechanisms adequate to ensure that outputs are comprehensible and amenable to appropriate intervention, Article 13 of the AI Act potentially incentivizes *interpretability by design* and *by default*, encouraging developers to integrate interpretability as a foundational architectural principle within high-risk systems, rather than relying on post-hoc explanatory supplements to opaque AI models⁵⁸. Such an interpretation, in my assessment, is undoubtedly the most protective approach, particularly in high-stakes domains such as law enforcement. At the same time, however, it risks constraining the

⁵³ The nature of the recipient – whether a public authority operating the system, the general public, a judicial body, or a technical expert – should be duly considered when providing explanations. V. ARYA, R.K. E. BELLAMY, P.-Y. CHEN, A. DHURANDHAR, M. HIND, S.C. HOFFMAN, S. HOUDE, Q.V. LIAO, R. LUSS, A. MOJSILOVIĆ, S. MOURAD, P. PEDEMONTE, R. RAGHAVENDRA, J. RICHARDS, P. SATTIGERI, K. SHANMUGAM, M. SINGH, K.R. VARSHNEY, D. WEI, Y. ZHANG. *One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques*, arXiv:1909.03012, 2019. Q.V. LIAO, K. R. VARSHNEY, *Human-Centered Explainable AI (XAI): From Algorithms to User Experiences*, arXiv:2110.10790, 2021.

⁵⁴ For a detailed discussion on this provision see GUTIÉRREZ DAVID, *Transparency and Provision of Information to Deployers in Article 13*, cited. Standards relating to transparency in AI systems are in the process of publication, notably the forthcoming ISO/IEC 12792 *Information technology — Artificial intelligence — Transparency taxonomy of AI systems*. Available at: <https://www.iso.org/standard/84111.html>.

⁵⁵ AI Act, Recital 72 “*High-risk AI systems, should be accompanied by appropriate information in the form of instructions of use*”. These instructions must be concise, complete, accurate, and clear, as well as relevant, accessible, and comprehensible to deployers. They must also cover all elements listed in Article 13(3) and include, at a minimum: the provider’s identity and contact details; the characteristics, capabilities, and performance limitations of the high-risk AI system; any predetermined system changes; oversight measures; the computational and hardware resources required, along with maintenance and update procedures; and where applicable, mechanisms for the collection and interpretation of logs. The information required under Article 13 mirrors, to some extent, that mandated by the GDPR, although it provides a higher level of specificity and detail.

⁵⁶ AI Act, Recital 72 “*High-risk AI systems should be designed in a manner to enable deployers to understand how the AI system works, evaluate its functionality, and comprehend its strengths and limitations*”. AI Act, Article 13(3)(b)(vii).

⁵⁷ AI Act, Article 14.

⁵⁸ See GUTIÉRREZ DAVID, *Transparency and Provision of Information to Deployers in Article 13*, cited, 689.

development and deployment of more complex models, which, despite their opacity, may in some circumstances offer valuable predictive capacity or uncover patterns that simpler, inherently interpretable models cannot as easily capture.

In cases where self-interpretable models are considered inadequate and reliance on uninterpretable models is required, the obligation to provide explanations becomes imperative⁵⁹. Explanations can be distinguished as global or local. While global explanations provide an overarching understanding of a model's behaviour, capturing general patterns and trends⁶⁰, local explanations focus on individual outputs, clarifying why a specific decision was reached in each case⁶¹. For instance, local explanation can show how much each feature contributed to the deviation of the model's prediction from the average risk score across all individuals. This elucidates both the determinative factors underlying the classification in a given instance and the manner in which these factors interact. Although post-hoc explanatory methods are identified as the sole solution to complex, inscrutable AI models, they have been heavily criticised as not being sufficiently reliable⁶².

Given its inherently technical nature, Article 13(1) must be anchored in detailed technical standards in order to be enforceable⁶³. As I argued in Chapter 3, this reliance on technical standards may be contentious because outsourcing the definition of transparency – and the threshold for what qualifies as “*sufficient transparency*” – to private standard-setting bodies effectively removes these fundamental determinations from the public domain. This issue becomes particularly critical in law enforcement, where public authorities frequently procure AI systems from private vendors, thereby shifting critical normative judgments from democratically accountable institutions to other entities whose incentives may not align with the public interest.

⁵⁹ Numerous techniques have been developed to explain the outputs of AI systems. One prominent approach is counterfactual explanations, which aim to illustrate how small changes in input could alter an automated decision. See S. WACHTER, B. MITTELSTADT, C. RUSSELL, *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, in *Harvard Journal of Law & Technology*, 31(2), 2018, 841-887. For a critical perspective on post-hoc explanation methods, see S. BORDT, M. FINCK, E. RAIDL, U. VON LUXBURG, *Post-Hoc Explanations Fail to Achieve their Purpose in Adversarial Contexts*, in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, Association for Computing Machinery, New York, NY, USA, 2022, 891–905. Available at: <https://doi.org/10.1145/3531146.3533153>.

⁶⁰ Techniques such as feature importance identify the variables that exert the greatest influence on a model's predictions, while rule extraction methods generate human-readable rules that approximate the decision-making processes of complex models. VARSHNEY, *Trustworthy Machine Learning*, cited, 163.

⁶¹ For examples of techniques used for local explanations – namely *Local Interpretable Model-agnostic Explanations* (LIME) and *Shapley Additive Explanations* (SHAP) – see the EDPS, *TechDispatch on Explainable Artificial Intelligence*, cited, 7-8. See also VARSHNEY, *Trustworthy Machine Learning*, cited, 163.

⁶² D. VALE, A. EL-SHARIF, M. ALI, *Explainable artificial intelligence (XAI) post-hoc explainability methods: Risks and limitations in non-discrimination law*, in *AI & Ethics*, 2(1), 2022, 1-12. Available at: <https://doi.org/10.1007/s43681-022-00142-y>. BORDT, FINCK, RAIDL, VON LUXBURG, *Post-Hoc Explanations Fail to Achieve their Purpose in Adversarial Contexts*, cited.

⁶³ For an overview of technical standards under development that are relevant to the transparency of AI systems, see GUTIÉRREZ DAVID, *Transparency and Provision of Information to Deployers in Article 13*, cited, 668. See also ISO/IEC DIS 12792:2024, currently under publication, available at <https://www.iso.org/standard/84111.html>.

2.4. *The fourth dimension of transparency: explainability*

Explainability refers to the extent to which a system can provide human-understandable justifications for its outputs, whether these are decisions, predictions, or recommendations⁶⁴. In the case of black-box models where supplementary explanation tools are needed, the concepts of interpretability and explainability may appear to overlap; however, this should not be the case, as these two dimensions operate at different levels: while interpretability is about understanding the inner working of the model, addressing the question of “how” the system arrived at a certain output (*i.e.*, the internal processes involved)⁶⁵, explainability is about making sense of the model’s behaviour, responding to the question of “why” the system produced such an output, explained in human terms. Explainability has therefore been defined as an “*active characteristic of AI models*”⁶⁶ which refers to their technical capacity to generate an explanation of their behaviour based on the data used, the results obtained, and the entire decision-making process. The literature specifically refers to explainability as the “*additional information*”⁶⁷ that can be obtained from the model to identify the reasons or rationale behind the outcome, making it meaningful to human users.

The AI Act refers to explainability on very few occasions, but no definitions can be found. An indirect reference to explainability is encapsulated in Article 13(3)(b)(iv) where it requires that the instruction for use include also “*the technical capabilities and characteristics of the high-risk AI system to provide information that is relevant to explain its output*”. However, in this case, it appears that the notion of explainability tends to overlap with that of interpretability, largely because the two concepts are not clearly distinguished within the AI Act⁶⁸. This lack of definitional clarity blurs the boundaries between what should be understood as the ability of a system to provide human-understandable reasons for its outputs (explainability) and the capacity to comprehend and analyse the internal mechanisms or decision-making processes of the system itself (interpretability). Nonetheless, these concepts are not equivalent and must not be treated as such⁶⁹. Because of this overlap, some scholars

⁶⁴ Explainability concentrates on “*providing clear and coherent explanations for specific model predictions or decisions*”. EDPS, *TechDispatch on Explainable Artificial Intelligence*, cited, 5. A. THAMPI, *Interpretable AI. Building explainable machine learning systems*, Manning, 2022, 14-15. In another case, explainability was defined as “*the ability of an AI system to relevantly communicate the reasoning behind its decision-making process*”. GYEVNAR, FERGUSON, SCHAFER, *Bridging the transparency gap: What can explainable AI learn from the AI Act?*, cited, 965. E. SPILLER, *Il diritto di comprendere, il dovere di spiegare. Explainability e intelligenza artificiale costituzionalmente orientata*, in *BioLaw Journal*, 2, 2021, 425.

⁶⁵ “*Interpretability of machine learning models is the aim to let people understand how the machine makes its predictions*”. VARSHNEY, *Trustworthy Machine Learning*, cited, 163.

⁶⁶ GUTIÉRREZ DAVID, *Transparency and provision of information to deployers in Article 13*, cited, 691.

⁶⁷ *Ibidem*, 692.

⁶⁸ Even within the *Ethics Guidelines on Trustworthy AI*, the two concepts were not clearly differentiated, suggesting that explainability “*concerns the ability to explain both the technical processes of an AI system and the related human decisions*”. HLEG, *Ethics Guidelines on Trustworthy AI*, cited, 18.

⁶⁹ In similar terms, cfr. GUTIÉRREZ DAVID, *Transparency and provision of information to deployers in Article 13*, cited, 694.

contend that the Regulation’s primary shortcoming lies in its limited treatment of explainability, whereas interpretability receives comparatively greater emphasis⁷⁰.

To this end, it may be helpful to read Article 13 in conjunction with Article 14⁷¹, which introduces the requirement of human oversight for high-risk AI systems. Following a by design approach, it mandates that systems be developed and configured so they can be effectively overseen by natural persons throughout their use. In fulfilment of the deployer’s obligation to ensure human oversight⁷², the system must be deployed in a manner that equips those assigned this role with the capacity to, among other things, correctly interpret its outputs, having regard to the interpretation tools and methods at their disposal⁷³. Applied to person-based predictive policing, this requirement entails that officers be provided with clear and accessible information concerning the relevant characteristics of the AI system, including, in particular, the training data underlying the model, thereby equipping them to understand the basis of a given risk score, scrutinise its reliability, and exercise their judgment to intervene, disregard, or override the system where circumstances so warrant.

Therefore, read together, these provisions may suggest that the AI Act effectively establishes an “explainability obligation”, aimed at ensuring that deployers are effectively able to interpret outputs generated by the AI system in the specific case. Where the output of an AI system serves to support decision-making in the sensitive domain of law enforcement, whether in connection with biometric identification or risk assessment, it is of paramount importance that the police officers concerned be in a position to understand the rationale underlying the system’s outputs. Such understanding is necessary in order to integrate such output with the police officers’ own professional judgment, experience, and available information when reaching a final decision⁷⁴. In any event, the availability of such interpretation tools is indispensable to deployers’ compliance with Article 26 of the AI Act, pursuant to which deployers are required to ensure human oversight and the ongoing monitoring of high-risk AI systems throughout their operation.

While Article 13 – potentially in conjunction with Article 14 – bound the provider, who shall adopt all the measures identified in the provision to enable the deployer to interpret the system’s output, such interpretation is in turn necessary to deliver an explanation of the overall decision-making

⁷⁰ SCHNEEBERGER ET AL., *The Tower of Babel in Explainable Artificial Intelligence (XAI)*, cited, 70.

⁷¹ Some authors argue that reading Article 13(1) in conjunction with Article 14(4)(c) implies that the Regulation imposes an “explainability obligation” on the provider. GUTIÉRREZ DAVID, *Transparency and provision of information to deployers in Article 13*, cited, 694.

⁷² AI Act, Article 26.

⁷³ AI Act, Article 14(4)(c).

⁷⁴ An interesting consideration arises in distinguishing between static and interactive explanations. While both appear to satisfy the interpretability and explainability requirements of the AI Act, interactive explanations may be better suited to ensuring effective human oversight. VARSHNEY, *Trustworthy Machine Learning*, cited, 167. “High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which they are in use”. AI Act, Article 14, paragraph 1.

process to the ultimate recipient: the affected individual⁷⁵. Individuals affected by an AI system generally want to understand and, where necessary, challenge its output, particularly when it leads to an adverse outcome for them. To this end, Article 86 includes a direct reference to explainability in the latter sense, introducing a right to explanation for affected individuals who are subject to a decision taken by a deployer based on the output of a high-risk AI system listed in Annex III, where such a decision produces legal effects or similarly significant impacts⁷⁶. Under Article 86 of the AI Act, deployers are obliged to provide to the affected individual all necessary information related to the role of the AI system within the decision-making process, as well as the principal factors underpinning the decision taken.

When an AI system relies on interpretable models, it is comparatively easier to provide affected individuals with the information necessary to understand and potentially contest decisions taken on the basis of outputs from high-risk AI systems listed in Annex III. By contrast, when such systems are built on non-interpretable models, ensuring the provision of all relevant information becomes significantly more challenging⁷⁷. Although the introduction of a right to explanation of individual decision-making shall be welcomed, as I will discuss in the subsequent section, the scope of application of this provision is very limited.

3. Considerations on the established legal framework with a focus on law enforcement

The discussion above makes it clear that the AI Act seeks to address transparency concerns by introducing a set of rules and obligations that approach the concept through its different dimensions⁷⁸. No single type of explanation suffices; rather, different types are needed depending on the purpose, the way the information must be conveyed, and the intended recipient⁷⁹. For instance, it may be necessary for decision-makers to possess a general model-level understanding of how the AI system operates to understand the predictions and combine them with their own information to make sound decisions. By contrast, the affected individual is likely to be primarily concerned with understanding

⁷⁵ Such explanations may also be of interest to data scientists, internal compliance officers, model validators, and external regulatory officials overseeing decision-making processes. VARSHNEY, *Trustworthy Machine Learning*, cited, 165.

⁷⁶ For a first discussion on the right to explanation see M.E. KAMINSKI, G. MALGIERI, *The Right to Explanation in the AI Act*, U of Colorado Law Legal Studies Research Paper No. 25-9, 2025. Available at SSRN: <https://ssrn.com/abstract=5194301>.

⁷⁷ MARCHETTI, TORCHIA, *AI and Public Administration*, in F. DECAROLIS, B. MARCHETTI, L. TORCHIA (Eds.), *The EU Digital Regulation and its Impact on Member States*, cited.

⁷⁸ The AI Act addresses: (i) the use of AI systems, irrespective of their risk category; (ii) log-keeping obligations; (iii) transparency regarding the system's internal functioning; and (iv) the provision of human-understandable explanations of the rationale underlying the system's outputs.

⁷⁹ VARSHNEY, *Trustworthy Machine Learning*, cited, 166-168.

which factors most decisively influenced the system's prediction, as such knowledge is indispensable for assessing the fairness of the decision, exercising the right to contestability, and ultimately safeguarding due process.

Despite the positive step of introducing transparency requirements and obligations in the AI Act, several important considerations remain to be addressed. First, transparency is a principle that, I would contend, ought to be safeguarded in all its dimensions, irrespective of the risk category assigned to an AI system. Notably, the European Parliament itself sought, through its proposed amendments, to accord transparency the status of a general principle within the AI Act⁸⁰, yet this provision was ultimately not retained in the final text. In any case, transparency obligations set out in Article 50 with regard to limited-risk AI systems are insufficient to address this gap, as they are too narrowly defined, addressing only the issue of communication and information disclosure, while leaving other aspects of transparency unregulated. Such a circumscribed approach raises particular concerns in the domain of law enforcement, where transparency in all its dimensions constitutes a precondition for both institutional accountability and preservation of public trust. A paradigmatic illustration is the deployment of place-based predictive policing tools which, as discussed in Chapter 3, are not even classified as limited-risk AI systems. Consequently, they remain entirely exempt from binding transparency obligations and are, at best, governed by non-compulsory and voluntary Codes of Conduct.

Second, transparency is a complex and multifaceted concept that, by its very nature, requires a plurality of both technical and regulatory responses. A major shortcoming of the AI Act lies in its failure to clearly distinguish between the different dimensions of transparency, each of which binds distinct actors in the AI value chain and targets different recipients⁸¹. For instance, Article 13 refers to both interpretability and explainability without explicitly differentiating between them, while Article 86 addresses explainability but shifts the recipient of information from the deployer to the affected individual. Similarly, communication and information obligations under Articles 26(11), 49, and 50 impose duties on different operators⁸² and designate different recipients⁸³.

⁸⁰ “‘transparency’ means that AI systems shall be developed and used in a way that allows appropriate traceability and explainability, while making humans aware that they communicate or interact with an AI system as well as duly informing users of the capabilities and limitations of that AI system and affected persons about their rights”. Amendment 213 Proposal for a regulation Article 4 a “General principles applicable to all AI systems”. Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).

⁸¹ M. BUSUIOC, D. CURTIN, M. ALMADA, *Reclaiming Transparency: Contesting the Logics of Secrecy within the AI Act*, 2 European Law Open 79, 2023, 91. Available at SSRN: <https://ssrn.com/abstract=4457949>.

⁸² Deployers for Articles 26(11), 49(3) and 50(3) and (4); providers for Articles 49(1) and (2) and 50(1) and (2).

⁸³ Affected individuals in the case of Articles 26(11) and 50, and the general public in the case of Article 49.

This lack of clarity is especially problematic in relation to a core principle of the AI Act. The absence of a coherent delineation of transparency's various dimensions, and of their interrelation, risks generating legal uncertainty and fragmented implementation, thereby undermining both the enforceability of the framework and the effective protection of public values and fundamental rights. A careful and consistent interpretation of these provisions is thus indispensable to ensure their effective application. It is further the responsibility of the European Commission to provide guidance on their practical implementation, with a view to ensuring coherence, uniformity, and legal certainty across Member States.

Another critique directed towards the AI Act, particularly prior to the introduction of Article 86, was that its approach to transparency was framed in predominantly technical terms and remained narrowly limited from a subjective, individual-centred perspective⁸⁴. The AI Act indeed directs the bulk of information obligations towards the deployer – typically the decision-maker employing the AI system in the decision-making process – rather than towards the affected individual, who usually remains a passive recipient of the outcome produced through the system. The introduction of Article 86 has, to some extent, improved this framework by recognising the position of the affected individual, in line with the “*right to an explanation*” enshrined in data protection legislation⁸⁵. However, the right to obtain an explanation within the AI Act is subject to important limitations⁸⁶. First, it applies only insofar as it is not already guaranteed under existing Union law, such as the GDPR or the LED⁸⁷.

⁸⁴ GUTIÉRREZ DAVID, *Transparency and provision of information to deployers in Article 13*, cited, 699.

⁸⁵ The so-called *right to an explanation* in data protection law has been primarily identified in connection with Article 22 of the GDPR. The body of literature on this topic is extensive. Among others, see B. GOODMAN, S. FLAXMAN, *European Union Regulations on Algorithmic Decision-Making and a “Right to Explanation”*, in *AI Magazine*, 38(3), 2017, 50-57. C. DIEFFAL, *The Normative Potential of the European Rule on Automated Decisions: A New Reading for Art. 22 GDPR*, in *Zeitschrift für Ausländisches Öffentliches Recht und Völkerrecht*, 80(4), 2020, 859-893. L. EDWARDS, M. VEALE, *Enslaving the Algorithm: From a “Right to an Explanation” to a “Right to Better Decisions”?*, in *IEEE Security & Privacy*, 17, 2018, 15; E. BAYAMLIOĞLU, *The Right to Contest Automated Decisions under the General Data Protection Regulation: Beyond the so-Called “Right to Explanation”*, in *Regulation & Governance*, 16(4), 2022, 1058-1078. E. FALLETTI, *Automated Decisions and Article 22 GDPR of the European Union: An Analysis of the Right to an “Explanation”*, in *Osservatorio dei diritti fondamentali in Europa*, 31, 2019, 1-15. L. METIKOŠ, J. AUSLOOS, *The Right to an Explanation in Practice: Insights from Case Law for the GDPR and the AI Act*, in *Law, Innovation and Technology*, March 2025, 1-36. Available at: <https://doi.org/10.1080/17579961.2025.2469349>. For a critical vision of the right to explanation see L. EDWARDS, M. VEALE, *Slave to the Algorithm? Why a ‘Right to an Explanation’ Is Probably Not the Remedy You Are Looking For*, in *Duke Law & Technology Review*, 16, 2017, 18-84.

⁸⁶ L. NANNINI, *Habemus a Right to an Explanation: so What? A Framework on Transparency-Explainability Functionality and Tensions in the EU AI Act*, in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 7(1), 2024, 1023-1035. The existence of a right to an explanation in the context of AI systems remains contested, as advanced machine learning models are often incapable of providing a comprehensive account of their internal functioning, being able to clarify only limited aspects of their decision-making processes. R. GUIDOTTI, A. MONREALE, S. RUGGIERI, F. TURINI, F. GIANNOTTI, D. PEDRESCHI, *A Survey Of Methods For Explaining Black Box Models*, in *ACM Computing Surveys*, Article No. 93, 51(5), 2018, 1-42. Available at: <https://doi.org/10.1145/3236009>. Accordingly, in cases involving the processing of personal data by means of fully automated decision-making, Article 22 GDPR—or, where relevant, Article 11 of the Law Enforcement Directive—may prevail over and supersede Article 86 of the AI Act. This interpretation is supported by Article 86(3), which explicitly states that “*This Article shall apply only to the extent that the right referred to in paragraph 1 is not otherwise provided for under Union law*”.

⁸⁷ In the context of law enforcement, for instance, see LED, Recital 38.

Second, the AI Act recognises this right exclusively for individuals affected by decisions taken through high-risk AI systems and, more specifically, those listed in Annex III. Therefore, it does not extend to high-risk systems that are instead regulated by harmonised legislation listed in Annex I. This limitation has a further, significant implication: AI systems that are not classified as high-risk fall entirely outside the scope of Article 86 and the corresponding individual right to an explanation. This is particularly relevant for place-based predictive policing tools which do not fall under the high-risk category. As a result, individuals living in or interacting with geographic areas flagged by such predictive tools – despite potentially experiencing increased police presence, stop-and-search practices, or other operational consequences – would not be entitled under the AI Act to understand the logic, input data categories, or reasoning behind the system’s risk predictions. This creates a notable regulatory gap: place-based predictive policing may significantly shape law enforcement deployment and affect fundamental rights, yet it remains outside the explanatory safeguards that the AI Act reserves for formally designated high-risk AI systems.

Moreover, in the specific context of law enforcement additional exceptions may apply, further narrowing the scope of Article 86. In such contexts, indeed, the right to an explanation must be carefully balanced against other legitimate interests, such as the protection of ongoing investigations or the confidentiality of sensitive operational activities (so-called “tactical barriers”)⁸⁸. For these reasons, Union or national law may provide for exceptions to, or restrictions on, the obligation to give explanations. In this author’s assessment, such limitations must be narrowly defined and applied with caution. Since these explanations are directed toward the affected individual – unlike the public-facing obligations under Article 49 – any restriction must remain proportionate, strictly necessary to achieve the intended objective, and accompanied by adequate safeguards to prevent abuse or overreach⁸⁹.

In any case, the practical application of such a right in AI-powered policing remains extremely challenging. In contexts involving AI-driven analytics (such as predictive policing) where an individual is designated as high-risk and thereby subjected to heightened surveillance, the decision that follows may not be directly predicated on the output of the AI system, a circumstance which may severely impede the ability of the individuals concerned to exercise their right to an explanation⁹⁰. In this respect, Article 26(11) may provide some support, as it requires deployers of high-risk AI systems listed in Annex III, when used to assist in decision-making concerning natural persons, to inform

⁸⁸ These interests are well established in Article 15 of the LED.

⁸⁹ In line with Article 15 of the LED.

⁹⁰ Some authors even argue that this Article does not apply in such contexts. In relation to AI-powered crime analysis, it has been observed that Article 86 is not applicable. A. SACHOULIDOU, *Harnessing AI for law enforcement: Solutions and boundaries from the forthcoming AI Act*, in *New Journal of European Criminal Law*, 2024, 15(2), 122.

those individuals that they are subject to the use of such systems, albeit subject to the limitations set out in Article 13 of the LED.

Notwithstanding the adoption of Article 86, scholarly commentary has observed that the AI Act's conceptualisation of transparency remains predominantly technocratic, being framed chiefly in terms of technical disclosure requirements and procedural formalities. Its orientation appears directed mainly towards enabling providers to demonstrate regulatory compliance, risking becoming only a box-ticking exercise, rather than toward securing the substantive, user-facing transparency necessary to ensure meaningful oversight and accountability in the operation of AI systems⁹¹.

This critique is closely linked to the one concerning Article 13(1), which has been criticised for its inability to guarantee robust transparency standards, owing to its dependence on vague and open-ended terminology. Indeed, it requires that AI systems be designed and developed in a manner that is “sufficiently transparent” and that “an appropriate type and degree of transparency” be ensured. It has been said that the absence of precise benchmarks confers considerable discretion upon providers in determining what qualifies as “sufficient” or “appropriate” transparency⁹². This, in turn, may heighten the risk of divergent interpretations and inconsistent levels of protection⁹³. Since the operationalisation of these requirements is, in practice, expected to be delegated to harmonised standards⁹⁴ and common specifications, the highlighted risk is shifting the balance toward compliance formalities at the expense of substantive transparency.

While these criticisms are not without merit and call for the devising of appropriate safeguards, it is this author's contention that the original intention of the European legislator may, in fact, have been of a different nature. Since the earliest soft law instruments, transparency has been identified as both a core objective of trustworthy AI⁹⁵ and one of the most difficult to realise. The complexity inherent in machine learning and deep learning inevitably constrains the practical attainability of this principle. Accordingly, the final version of the provision – differing from the original version – does not pursue

⁹¹ GUTIÉRREZ DAVID, *Transparency and provision of information to deployers in Article 13*, cited, 699-700. D. ONITIU, *The limits of explainability & human oversight in the EU Commission's proposal for the Regulation on AI - a critical approach focusing on medical diagnostic systems*, in *Information & Communications Technology Law*, 32(2), 2023, 174; D. SCHNEEBERGER ET AL., *The Tower of Babel in Explainable Artificial Intelligence (XAI)*, cited, 65 and 70; P. HACKER, J.H. PASSOTH, *Varieties of AI Explanations Under the Law. From the GDPR to the AIA, and Beyond*, in A. HOLZINGER, R. GOEBEL, R. FONG, T. MOON, K.R. MÜLLER, W. SAMEK (Eds.) *xxAI - Beyond Explainable AI. xxAI 2020, Lecture Notes in Computer Science*, 13200. Springer, 2022, 359.

⁹² M. BUSUIOC, D. CURTIN, M. ALMADA, *Reclaiming Transparency: Contesting the Logics of Secrecy within the AI Act*, cited, 98.

⁹³ To avoid divergent interpretations and applications, explainability metrics have been proposed as a possible reference benchmark. See F. SOVRANO, S. SAPIENZA, M. PALMIRANI, F. VITALI, *Metrics, Explainability and the European AI Act Proposal*, in *J*, 5(1), 2022, 126-138.

⁹⁴ The European Commission has already issued a standardisation request. European Commission, Commission Implementing Decision on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence, C(2023)3215, 22 May 2023.

⁹⁵ See HLEG, *Ethics Guidelines on Trustworthy AI*, cited.

full transparency but instead requires a “sufficient” or “adequate” degree thereof. While such indeterminate terms may appear vague, they acknowledge the technological limits that preclude absolute transparency. Indeed, demanding full transparency would either exclude the use of inherently opaque systems or reduce disclosure to unintelligible numerical data, thereby rendering transparency a merely formal safeguard. Therefore, a plausible interpretation of these vague formulations is that the effectiveness of transparency requirements depends on calibrating both the type and the level of information disclosed to the characteristics of the intended recipients, taking into account their background, knowledge, and expertise, as well as the purposes for which the explanation is sought⁹⁶.

The use of such indeterminate terminology may thus permit a substantively effective form of transparency, calibrated to the operator within the deployer’s organisational structure who is responsible for implementing the system. Only in this way can the deployer adequately fulfil its obligations. These include, *inter alia*, the requirement to ensure effective human oversight by natural persons; to guarantee that input data – where the deployer exercises control over them – are relevant and sufficiently representative for the intended purpose of the system; to continuously monitor the operation of the system in accordance with the instructions for use; to retain automatically generated logs under their control for the period defined in Article 26(6) in order to inform the provider and, where requested, the market surveillance authority pursuant to Articles 72, 73, and 79. Consequently, while providers formally retain “discretion” over transparency choices, they must structure them in a way that ensures an appropriate quality and degree of transparency, considering the identity and functional role of the intended recipient of the information⁹⁷. Only in such circumstances can Article 86 be meaningfully applied, as the deployer would be in a position to furnish the affected individual with sufficient information concerning the AI system’s output, on the basis of the information received from the provider⁹⁸.

This process is inherently complex, and a significant risk lies in providers overloading recipients with information that is excessively technical, thereby undermining its comprehensibility and practical utility. Close cooperation between the provider and the deployer should therefore be actively promoted and incentivised. The decision-maker – namely, the deployer of the AI system and the ‘problem owner’ – shall play an active role throughout the entire AI lifecycle⁹⁹ and such involvement is primarily to be secured through contractual negotiations. In addition, as reliance on harmonised standards appears inevitable, they can serve as a means of ensuring appropriate differentiation and

⁹⁶ In these terms, VARSHNEY, *Trustworthy Machine Learning*, cited, 164.

⁹⁷ Of course, this does not diminish the deployer’s obligation to ensure “a sufficient level of AI literacy of their staff and other persons dealing with the operation and use of AI systems on their behalf”.

⁹⁸ OLIVATO, *The AI Act: regulatory and technical challenges*, cited, 264.

⁹⁹ VARSHNEY, *Trustworthy Machine Learning*, cited, 14.

facilitating the dissemination of sound industry practices. For instance, the IEEE *Standard for Transparency of Autonomous Systems*¹⁰⁰ expressly identifies three categories of users – non-expert users¹⁰¹, domain-expert users¹⁰², and super-users¹⁰³ – and prescribes differentiated levels of transparency accordingly¹⁰⁴. At the same time, national authorities must be adequately equipped to detect potential malpractices and to intervene effectively to prevent their continuation.

As discussed above, while some AI models are interpretable by their very nature¹⁰⁵, and therefore do not pose particular problems in their understandability, some others are uninterpretable. The latter models are often referred to as “opaque”¹⁰⁶ due to their inherent complexity and their incapacity to provide direct explanations of how a given output has been generated. Their deployment has led to what is commonly referred to as the “black-box problem”¹⁰⁷, which constitutes a form of technical barrier¹⁰⁸ to transparency.

Although less interpretable models (such as deep neural networks) are often noted to achieve higher performance than more interpretable ones¹⁰⁹, research in the field of Explainable AI seeks to develop techniques that mitigate this trade-off, aiming to combine high performance with greater interpretability¹¹⁰. Nonetheless, several scholars have challenged the assumption of an inevitable trade-off between accuracy and interpretability, with Cynthia Rudin being the most prominent

¹⁰⁰ *IEEE Standard for Transparency of Autonomous Systems*, IEEE 7001-2021. Within the standards, three categories of users are identified — non-expert users, domain expert users, and super-users — along with five corresponding levels of transparency. A first example of this approach can be found in the *UK Algorithmic Transparency Recording Standard*, which requires public administrations to provide information at two distinct levels. Available at: <https://www.gov.uk/government/collections/algorithmic-transparency-recording-standard-hub>.

¹⁰¹ “Persons who have only a brief interaction or who interact every day with an autonomous system”. IEEE 7001-2021, paragraph 3.1.

¹⁰² “Persons who carry some responsibility for how an autonomous system is used or are responsible for operating and supervising autonomous systems”. IEEE 7001-2021, paragraph 3.1.

¹⁰³ “Experts not only in autonomous systems but also in the particular systems for which they are responsible”. IEEE 7001-2021, paragraph 3.1.

¹⁰⁴ Transparency is defined as “A transfer of information from an autonomous system or its designers to a stakeholder that is truthful; contains information relevant to the causes of some action, decision, or behavior; and is presented at a level of abstraction and in a form meaningful to the stakeholder. Transparency should be mindful of the stakeholders’ likely perception and comprehension, and should avoid disclosing information in a manner that, while technically true, is framed in a way that leads to misapprehension”. IEEE 7001-2021.

¹⁰⁵ VARSHNEY, *Trustworthy Machine Learning*, cited, 169. Examples of directly interpretable models, because of their relatively simple form, are Decision trees, Boolean rule sets, logistic regression, GAMs, GLRMs.

¹⁰⁶ See J. BURRELL, *How the machine ‘thinks’: Understanding opacity in machine learning algorithms*, in *Big Data & Society*, 3(1), 2016, 1-12.

¹⁰⁷ GUIDOTTI, MONREALE, RUGGIERI, TURINI, GIANNOTTI, PEDRESCHI, *A Survey Of Methods For Explaining Black Box Models*, cited, 5; A. ADADI, M. BERRADA, *Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)*, in *IEEE Access*, 6, 2018, 52138-52160. Available at: <https://ieeexplore.ieee.org/document/8466590>. One of the first and relevant works on the topic is F. PASQUALE, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, 2015.

¹⁰⁸ A.G. FERGUSON, *Illuminating Black Data Policing*, in *15 Ohio State Journal of Criminal Law* 503, 2018. Available at: SSRN: <https://ssrn.com/abstract=3142263>.

¹⁰⁹ A. BELL, I. SOLANO-KAMAICO, O. NOV, J. STOYANOVICH, *It’s Just Not That Simple: An Empirical Study of the Accuracy-Explainability Trade-off in Machine Learning for Public Policy*, in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT ‘22)*, 2022, Association for Computing Machinery, New York, NY, USA, 248–266. Available at: <https://doi.org/10.1145/3531146.3533090>.

¹¹⁰ D. GUNNING, D.W. AHA, *DARPA’s Explainable Artificial Intelligence Program*, in *AI MAGAZINE*, 2019, 44-58.

proponent of this critique. Rudin argues that inherently interpretable models – when working with structured data and representative and relevant attributes – can achieve levels of predictive performance comparable to those of black-box models, thereby undermining the rationale for privileging opacity in high-stakes contexts, such as criminal justice, medicine and, by extension, law enforcement¹¹¹. To illustrate this, she makes the examples of models derived from specialised algorithms¹¹² which achieve predictive accuracy comparable to that of the proprietary COMPAS model, despite being far more transparent and readily interpretable. Rudin therefore believes that the current practice of using complex black-box models and then relying on post-hoc explanations should be replaced by the development and adoption of models that are inherently interpretable, especially in high-stakes areas¹¹³.

Since the use of uninterpretable models inevitably amplifies the risks of errors, biases, and potential rights violations¹¹⁴ and given the particularly sensitive nature of the law enforcement sector, the adoption of inherently interpretable models, as advocated by Rudin, appears to offer stronger and more reliable safeguards for the protection of individuals. However, this solution may not always be practicable, as in certain cases the deployment of more complex models may be required, for instance to achieve the desired level of performance or accuracy. Nevertheless, Professor Richard A. Berk observes that there are several ways to enhance the interpretability of machine learning models. First, models can be constructed in ways that are inherently easier to understand¹¹⁵. Second, various tools and techniques have been developed to improve the understanding of complex models, often through visualizations that illustrate which predictors contribute most to the outcomes¹¹⁶. Third, it is frequently possible to address specific questions about what occurs ‘under the hood’ of the model¹¹⁷.

¹¹¹ C. RUDIN, *Stop Explaining Black-Box Machine Learning Models for High-Stakes Decisions and Use Interpretable Models Instead*, in *Nature Machine Intelligence*, 1, 2019, 206–215.

¹¹² Risk-Supersparse Linear Integer Models (RiskSLIM) and Certifiably Optimal Rule Lists (CORELS).

¹¹³ Post-hoc explanations may be misleading for several reasons: they do not necessarily reflect the model’s actual computations; they often omit key details needed to understand its internal logic; they may not align with contexts requiring integration of external information; and they can generate unnecessarily complex reasoning chains, heightening the risk of human error. The Information Commissioner’s Office and the Alan Turing Institute, in *Explaining Decisions Made with AI*, cited, 68–69, have likewise underscored the importance of favouring interpretable models over black-box systems in high-stakes domains.

¹¹⁴ C. COGLIANESE, A. LAI, *Assessing Automated Administration*, U of Penn Law School, Public Law Research Paper No 22-28, in J.B. BULLOCK ET AL. (Eds.), *Oxford Handbook on AI Governance*, 2022. Available at SSRN: <https://ssrn.com/abstract=4084844>.

¹¹⁵ J. ZENG, B. USTUN, C. RUDIN, *Interpretable classification models for recidivism prediction*, in *Journal of the Royal Statistical Society*, 180(3), 2017, 689–722.

¹¹⁶ Berk provides two illustrative examples of methods for assessing variable importance. The first method tests how much model accuracy decreases when each predictor is removed in turn; if accuracy drops substantially, the predictor is deemed important. He cites L. BREIMAN, *Random Forests*, in *Machine Learning*, 45, 2001, 5–32. The second method examines how each predictor relates to the outcome while holding all other predictors constant, but without relying on the usual adjustments used in linear regression. For this, Berk refers to T. HASTIE, R. TIBSHIRANI, J. FRIEDMAN, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2009.

¹¹⁷ “For example, one can take a single case and obtain the fitted class. Then for the same case, the value of some predictor such as age is changed. One can learn whether the fitted class would have been meaningfully different if the individual were 10 years older, with everything else about the individual fixed. In this manner, the way the algorithm uses age can

Finally, when the code is available it specifies in detail how risk scores are computed. From these considerations, one may conclude that within law enforcement contexts, AI systems must either employ inherently interpretable models or, where such approaches prove infeasible, black-box architectures must be supplemented with robust explainability mechanisms to render their outputs comprehensible to human operators¹¹⁸.

In the following subsections, I will analyse the relevant provisions of the AI Act with reference to the transparency barriers outlined in Chapter 2. As discussed in that Chapter, Andrew Ferguson – through a notable contribution – introduced the concept of transparency barriers to describe the various obstacles that inevitably limit full transparency in relation to AI systems. These barriers arise from technical constraints, legal limitations, and practical or operational considerations, all of which must be adequately taken into account when assessing the transparency regime established by the AI Act.

3.1. Technical barriers

As discussed above, Article 13(1) appears to discourage the proliferation of so-called black-box models by promoting interpretability *by design* and *by default* as guiding principles¹¹⁹. Where the development of inherently interpretable models is not feasible, the provision requires reliance on complementary measures and tools aimed at rendering the model sufficiently understandable to relevant stakeholders. This reflects an acknowledgment that opacity cannot always be eliminated at the design stage but must at least be compensated through mechanisms capable of ensuring an adequate degree of intelligibility¹²⁰. Recital 72 provides important insights in this regard. To assist deployers in fulfilling their obligations under the Regulation – namely those set out in Article 26 – the Recital underscores the centrality of transparency, understood in both its “interpretability” and

be easily demonstrated”. R.A. BERK, *Artificial Intelligence, Predictive Policing, and Risk Assessment for Law Enforcement*, in *Annual Review of Criminology*, 4, 2021, 209–237.

¹¹⁸ A line of research known as Explainable AI (XAI) focuses on developing approximation techniques – namely secondary, simplified models – that replicate, to a significant extent, the behaviour of black-box models by elucidating the relationships between input features and predictive outcomes. For critical considerations on such approximation methods, see C. RUDIN, *Stop Explaining Black-Box Machine Learning Models for High-Stakes Decisions and Use Interpretable Models Instead*, cited. See also S. WACHTER ET AL., *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*, in *Harvard Journal of Law and Technology*, 31(2), 2018, 850. For a discussion on explainable AI in the criminal sector see X. LI, *Explainable AI (XAI) Empowering Criminal Psychological Profiling: In Pursuit of Transparency and Trust within the Black Box*, in *International Journal of Science*, 12(7), 2025, 167-175. VARSHNEY, *Trustworthy Machine Learning*, cited, 175.

¹¹⁹ As some authors point out, “it is not clear, in any case, whether this was the legislator’s intention, since such an approach could stifle innovation”. GUTIÉRREZ DAVID, *Transparency and provision of information to deployers in Article 13*, cited, 714.

¹²⁰ As discussed above, the AI Act, however, does not specify the type or degree of transparency required to ensure compliance with its provisions.

“explainability” dimensions, as a prerequisite for high-risk AI systems, prior to their placement on the market or their being put into service.

Moreover, it specifies that such systems must be accompanied by adequate documentation and instructions for use. Indeed, deployers are required by Article 26(1) to adopt both technical and organisational measures that are adequate to ensure that the AI system is used strictly in accordance with the instructions given by the provider¹²¹. To this end, providers shall furnish such instructions in a way that ensure they are “*concise, complete, correct and clear, as well as relevant, accessible and comprehensible to deployers*”. Indeed, these instructions serve as the primary channel through which deployers understand the intended purpose, capabilities, and limitations of the AI system in order to make informed decisions regarding its operation¹²². The information presented in the instructions for use are briefly described in the figure below.

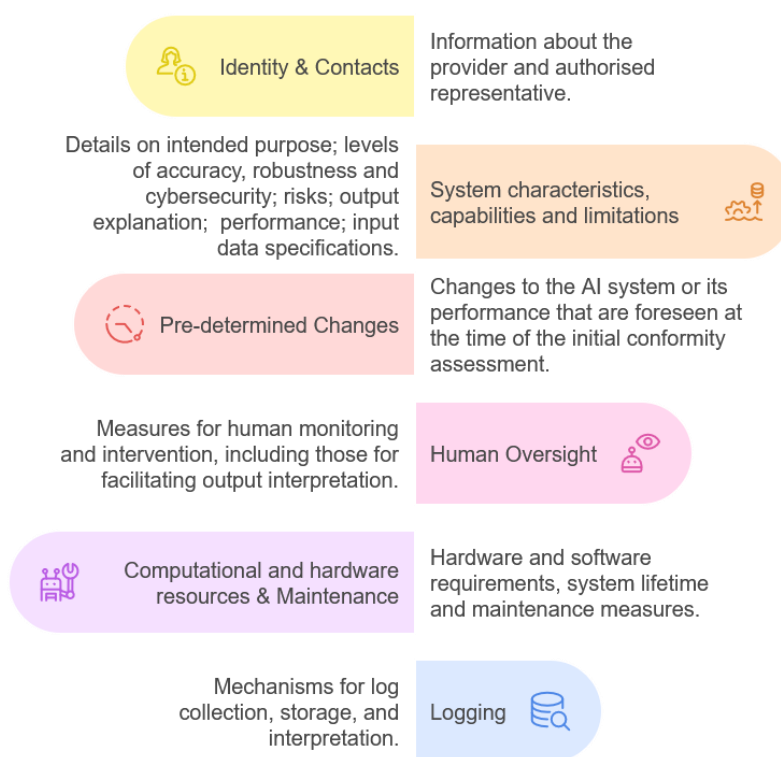


Figure 12: Article 13 – Information in the instructions for use. Own elaboration with Napkin.

An interesting aspect of instructions for use is that they are not limited to the technical aspects of AI systems, but they cover broader socio-technical considerations. Regarding performance metrics,

¹²¹ AI Act, Article 26(1).

¹²² “Deployers should, *inter alia*, be in a better position to make the correct choice of the system that they intend to use in light of the obligations applicable to them, be educated about the intended and precluded uses, and use the AI system correctly and as appropriate”. For a discussion on the instructions for use see GUTIÉRREZ DAVID, *Transparency and provision of information to deployers in Article 13*, cited, 704-713.

Article 13 mandates disclosure of information concerning the system's performance with respect to the specific demographic groups or individuals for whom the system is designed to be utilized. As examined in Chapter 2, the accuracy of AI systems may vary substantially across different demographic populations upon which they are deployed. This variance frequently derives from insufficient representativeness within training datasets, which may fail to adequately reflect the diversity of the population to which the system is subsequently applied. Consequently, certain groups may experience systematic disadvantage, as algorithmic performance tends to deteriorate when applied to individuals or communities that were underrepresented or inadequately represented during the training phase. Correspondingly, the instructions for use must specify measures for human monitoring and intervention, including provisions to facilitate the interpretation of system outputs, thereby ensuring effective human oversight mechanisms¹²³.

This information reflects the broader objective of ensuring that deployers are not merely passive users of complex technologies but are instead equipped with the necessary knowledge to understand the system's functioning, assess its limitations, and integrate its outputs responsibly within their decision-making processes¹²⁴. Moreover, deployers of high-risk AI systems are required to actively monitor the operation and performance of such systems in accordance with the instructions provided by the system's provider. This monitoring obligation is not passive and deployers must ensure that the system is functioning as intended, remains within the limits of safe and lawful use, and does not create unforeseen risks. If, during the course of deployment, the system presents or is likely to present a risk to health, safety, fundamental rights, or any other interest protected under Article 79(1) of the AI Act¹²⁵, the deployer has a duty to promptly notify the provider, so that it can take corrective actions, update risk assessments, or issue revised instructions, thereby maintaining compliance with legal and safety requirements. Essentially, Article 26(5) identifies two linked obligations: first, deployers must continuously monitor the system based on provider instructions; second, deployers must inform the provider without delay if risks emerge during use.

Although the introduction of an obligation to provide information on high-risk AI systems may generally be regarded as a positive measure, some scholars have expressed concerns that such duties could, rather than enhancing users' awareness and control, lead to an overload of information, thereby

¹²³ AI Act, Articles 14 and 26.

¹²⁴ GUTIÉRREZ DAVID, *Transparency and provision of information to deployers in Article 13*, cited, 714. The information should be presented in different ways, depending on the stakeholder who is intended to read it. J. GERLINGS, M. S. JENSEN, A. SHOLLO, *Explainable AI, but explainable to whom?*, 2021, arXiv:2106.05568.

¹²⁵ AI Act, Article 79(1): "AI systems presenting a risk shall be understood as a "product presenting a risk" as defined in Article 3, point 19 of Regulation (EU) 2019/1020, in so far as they present risks to the health or safety, or to fundamental rights, of persons".

creating an appearance of transparency without ensuring its effective realisation¹²⁶. This objection, however, appears less persuasive in the context considered in this Thesis. Deployers are expected to acquire a sufficient level of knowledge and capacity to read this information and ensure the proper and responsible use of the AI system¹²⁷, since they are professional users¹²⁸. More problematic is the structural asymmetry arising from the fact that information described above is generally held exclusively by the provider and necessarily mediated through them, with strong vested interests likely to shape both the scope and the framing of the disclosure¹²⁹. This consideration introduces another category of barriers identified in Chapter 2: technological barriers.

3.2. Technological barriers

Technological barriers encompass the limitations imposed by AI providers upon whom law enforcement authorities customarily depend¹³⁰. As discussed in Chapter 2, law enforcement authorities rarely develop such systems internally¹³¹; rather, they predominantly rely on private sector entities for the technologies they deploy¹³². This reliance creates a structural asymmetry between providers and deployers, as access to crucial information about the system's design, functioning, and datasets is often restricted by intellectual property rights opposed by private companies¹³³. Vendors may invoke copyright and trade secret law to shield source code, model architectures, or training data from disclosure, thereby preventing law enforcement authorities – and deployers in general – from

¹²⁶ C. NORVAL, K. CORNELIUS, J. COBBE, J. SINGH, *Disclosure by Design: Designing information disclosures to support meaningful transparency and accountability*, in 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), 21–24 June 2022, Seoul, Republic of Korea, ACM, New York, NY, USA, 2022, , 679-690.

¹²⁷ AI Act, Article 4; Recital 20. “Providers and deployers of AI systems shall take measures to ensure, to their best extent, a sufficient level of AI literacy of their staff and other persons dealing with the operation and use of AI systems on their behalf, taking into account their technical knowledge, experience, education and training and the context the AI systems are to be used in, and considering the persons or groups of persons on whom the AI systems are to be used”. Therefore, all the necessary information enabling stakeholders to make informed decisions regarding AI systems shall be provided. “Those notions may vary with regard to the relevant context and can include understanding the correct application of technical elements during the AI system's development phase, the measures to be applied during its use, the suitable ways in which to interpret the AI system's output, and, in the case of affected persons, the knowledge necessary to understand how decisions taken with the assistance of AI will have an impact on them”. Cfr. European Commission, FAQs, AI Literacy - Questions & Answers. Available at: <https://digital-strategy.ec.europa.eu/en/faqs/ai-literacy-questions-answers>.

¹²⁸ A.P. ORTIGOSA, *General regime of obligations for providers and deployers in the Artificial Intelligence Act*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 475.

¹²⁹ BUSUIOC, CURTIN, ALMADA, *Reclaiming Transparency: Contesting the Logics of Secrecy within the AI Act*, cited, 89.

¹³⁰ FERGUSON, *Illuminating Black Data Policing*, cited, 13.

¹³¹ For general considerations on whether to develop the system in-house or to outsource it, see Ada Lovelace Institute, *Buying AI. Is the public sector equipped to procure technology in the public interest?*, September 2024. Available at: <https://www.adalovelaceinstitute.org/report/buying-ai-procurement/>.

¹³² See S. BRAYNE, *Predict and Surveil: Data, Discretion, and the Future of Policing*, Oxford University Press, 2020.

¹³³ For a discussion on this topic see M. MAGGIOLINO, *EU Trade Secrets Law and Algorithmic Transparency*, Bocconi Legal Studies Research Paper No. 3363178, 2019. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3363178.

fully understanding or scrutinizing the tools they deploy¹³⁴. While such protections serve to safeguard innovation and the economic interests of providers¹³⁵, they simultaneously generate a barrier to transparency which is particularly critical for the use of AI in the public sector.

To such concerns, it may be replied that the protection of undisclosed know-how and business information is widely recognised as being subject to limitations where overriding public interests are at stake¹³⁶. Much has been written about the disclosure of the source code of software used by public authorities. It is beyond doubt, in my analysis, that public administrations should have access to the source code of the software they use. As Giovanni Sartor maintains, access to the source code is essential for a comprehensive understanding of the software's functioning, extending beyond the mere observation of its externally perceptible behaviour¹³⁷ and it is indispensable for intervening in the software, whether to correct existing errors or to modify it in order to adapt to new situations.

However, several courts across Europe have gone further, addressing this issue in cases concerning applicants' right of access to the source code of software used by public administrations in carrying out their functions¹³⁸. In Italy¹³⁹, for instance, the judiciary has emphasised that, when public authorities rely on software to perform activities of public relevance, the software becomes a direct expression of the activity carried out by the public administration¹⁴⁰ and therefore access to the source

¹³⁴ This inevitably adds to the opacity of AI systems, as discussed above.

¹³⁵ S. SCHROFF, *The purpose of copyright – moving beyond the theory*, in *Journal of Intellectual Property Law & Practice*, 16(11), 2021, 1262–1272. P. HUTUKKA, *Copyright Law in the European Union, the United States and China*, in *IIC - International Review of Intellectual Property and Competition Law*, 54, 2023, 1044–1080.

¹³⁶ TAR Lazio-Roma, Judgement of 30 June 2021, No. 7769. The case originated when candidates in a public selection procedure requested access to the source code of the software used for the written examination. After a partial disclosure of the source code, the candidates brought an action against the public administration, seeking full access. The administration refused to grant such access, but the court ultimately ordered it to allow the candidates full access to the source code. In legal scholarship, however, well-founded arguments have been made by those who contend that access to the source code of AI systems may not, in itself, be sufficient to understand the mechanisms underlying such technology. T. SCANTAMBURLO, A. CHARLESWORTH, N. CRISTIANINI, *Machine Decisions and Human Consequences*, in K. YEUNG, M. LODGE (Eds.), *Algorithmic Regulation*, Oxford, 2019, 71. Article 1(2)(b) of Directive 2016/943 on the protection of undisclosed know-how and business information (trade secrets) states that such protection “does not affect the application of Union or national rules requiring trade secrets to be disclosed to the public or to administrative or judicial authorities for the performance of their duties”.

¹³⁷ G. SARTOR, *L'informatica giuridica e le tecnologie dell'informazione*, Giappichelli, 2022, 110.

¹³⁸ Apart from Court's cases, the disclosure of the source code seems to be accepted as a best practice, as it emerges from Etalab, *Overview of the guide*. Available at: <https://guide-juridique-logiciel-libre.etalab.gouv.fr/>.

¹³⁹ For a discussion on digitalisation and use of AI by public administrations in Italy see M. MERLER, C. RAMOTTI, *Italian public administration in the face of digital transformation*, in F. DECAROLIS, B. MARCHETTI, L. TORCHIA (Eds.), *The EU Digital Regulation and its Impact on Member States*, cited, 35-76. Italian Administrative Courts clarified that the principle of transparency requires that, in algorithmic administrative decisions, the automated process be made not only knowable in its technical aspects but also understandable, through an explanation that translates it into the “legal rule” underlying it, thus rendering it intelligible to its addressees. This is intended, on the one hand, to ensure the full exercise of the right of defence by the individual affected by the measure, pursuant to Articles 24 and 113 of the Italian Constitution, and, on the other, to enable full judicial review of legality by administrative judges. Among the most relevant judgements see Italian Council of State, Section VI, Judgement of 4 February 2020, No. 881; Italian Council of State, Section VI, Judgement of 13 December 2019, No. 8472 and 8473; Italian Council of State, Section VI, Judgement of 8 April 2019, No. 2270; TAR Lazio-Roma, Sez. III-bis, Judgement of 19 April 2019, No. 5139.

¹⁴⁰ In TAR Lazio-Rome, Judgment No. 7769/2021, the Court reaffirmed that “source codes” qualify as documents accessible under Law No. 241/1990, given their substantive administrative value, irrespective of their origin. Similarly, TAR Lazio-Roma, sez. III-bis, 30 June 2020, No. 7370.

code shall be given¹⁴¹. In light of this, intellectual rights protection cannot be invoked in a manner that undermines citizens' right to scrutinise administrative decisions and, where necessary, to challenge them¹⁴².

Similar conclusions have been reached by the Spanish Tribunal Supremo in a ruling of September 2025¹⁴³. Following the unfavourable rulings in the lower instance¹⁴⁴, the Fundación Ciudadana Civio lodged a cassation appeal before the Tribunal Supremo, claiming its right to access the source code of the BOSCO application, a software used by the Ministry for the Ecological Transition. In this ruling, the Tribunal Supremo ruled not only that citizens have the right to access the source code of government algorithms used in automated decision-making processes but also articulated several important legal principles¹⁴⁵. First, the right of access to public information transcends its merely administrative dimension and attains the rank of a constitutional right. Second, even where public authorities resort to automated decision-making, transparency remains indispensable to guarantee that citizens possess the necessary information to understand and oversee the decision-making process (so-called "democratic governance"). Third, limitations on access to public information are admissible, but they must be duly justified and proportionate in light of the specific circumstances of each case. Fourth, where algorithmic decision-making bears upon the enjoyment of social rights, the public interest in transparency will, as a rule, prevail over administrative concerns related to

¹⁴¹ An exception to such accessibility is admissible only in specific cases of a public nature expressly provided for under the rules governing access to "administrative documents" held by the public administration (Law No. 241/1990, Article 24). Among these cases, access is excluded where the documents concern the structures, means, equipment, personnel, and actions strictly instrumental to the protection of public order or to the prevention and repression of crime. Such interests are contemplated by Article 24 L. 241/1990; Article 8, letter c) D.P.R. 27 June 1992, No. 352; Article 3 Decree 10 May 1994, n. 415. For an interesting discussion on the access to algorithms in the American context, see H. BLOCH-WEBBA, *Articles. Access To Algorithms*, in *Fordham Law Review*, 2020, 88(4), 1265-1314.

¹⁴² In a previous case, the Italian Council of State expressly held that the alleged confidentiality claimed by companies producing software used by public administrations is irrelevant, since, by making such tools available for the exercise of public authority, they must accept the resulting obligations of transparency (Italian Council of State, Section IV, 4 February 2020, No. 881, para. 10). Moreover, the Court considered that any requirements related to confidentiality and the protection of industrial know-how are to be regarded as secondary, precisely because of the countervailing principles of sharing, accessibility, and software reuse that underpin the Code of Digital Administration.

¹⁴³ Tribunal Supremo, Sala de lo Contencioso-Administrativo Sección Tercera, Sentencia núm. 1119/2025, 11 September 2025. Available at: <https://vlex.es/vid/1091644276>.

¹⁴⁴ In September 2018, Fundación Ciudadana Civio requested access, through the transparency portal, to the technical documentation, test results, and source code of BOSCO. BOSCO is a telematic system used to verify whether applicants meet the conditions for the social electricity bonus, that is, discounted tariff for vulnerable consumers. The request was based on the need to understand how it processes eligibility and to verify that its decision logic complies with the applicable legal framework. In February 2019, the Council for Transparency and Good Governance (CTBG) partially upheld the request: it ordered the disclosure of the technical specifications and test results but denied access to the source code, citing intellectual property rights and security concerns. In December 2021, the Central Administrative Court dismissed Fundación Ciudadana Civio's appeal, thereby upholding the denial of access. In April 2024, the National Court confirmed that decision on appeal. Subsequently, Fundación Ciudadana Civio filed a cassation appeal before the Supreme Court.

¹⁴⁵ The Government opposed Fundación Ciudadana Civio's request by invoking intellectual property rights, the potential exposure of the system to cyberattacks, with inevitable consequences for data security, and concerns related to the protection of personal data. Conversely, Fundación Ciudadana Civio based its appeal on the constitutional right of access to public information, arguing that such access enables algorithmic accountability, error detection, and democratic oversight.

cybersecurity¹⁴⁶ and intellectual property protection. In particular, while intellectual property rights provide necessary protection for software, such protection cannot operate as an automatic barrier to transparency, and a proper balance must therefore be struck between these two interests.

In comparable terms, the French Conseil constitutionnel has established that transparency requirements supersede trade secret and intellectual property claims. Accordingly, where the algorithm's mechanisms cannot be disclosed due to proprietary protections, public authorities are precluded from rendering individual decision "*on the exclusive basis of this algorithm*"¹⁴⁷.

In Italy, these principles are reaffirmed in Article 30 of the Italian Code of Public Procurement, which stipulates that, when acquiring or developing technological solutions (including AI) contracting authorities and granting entities must guarantee the *availability* of the source code, the accompanying documentation and all other elements necessary for understanding the operating logic of the system¹⁴⁸. Such a provision inevitably affects the procurement process, requiring public administrations, during negotiations with private entities, to ensure the disclosure of a wide range of crucial information¹⁴⁹. In a subsequent decision, however, the Council of State clarified that Article 30(2)(a), applies exclusively to so-called "administrative algorithmic decisions", namely cases in which the software itself substantially makes the decision¹⁵⁰. By contrast, in all cases where "algorithms" (or, more precisely, software) are merely used to support human decision-making (decision-support systems), such provision does not apply. This is because, in the Court's reasoning, software that merely supports decision-making – while responsibility remains entirely with the human operator – falls within the traditional approach, which regards information and communication technologies as a mere aid to the performance of administrative activity in its classic operational modalities¹⁵¹.

In this author's contention, however, these conclusions are susceptible to critique, at least in part. First, the idea that an "algorithm" that merely supports decisions ultimately entrusted to human judgement remains within the traditional understanding of information technology as a simple aid to

¹⁴⁶ The Tribunal Supremo acknowledged the existence of cybersecurity risks but held that security concerns must be specific and proportionate, rather than generic. The Court further suggested that confidentiality agreements and usage restrictions may constitute appropriate safeguards, instead of a complete denial of access.

¹⁴⁷ "[W]hen the principles of the functioning of an algorithm cannot be communicated without infringing on one of the secrets or interests set out under Section 2° of Article L. 311-5 of the Code of the Relationship Between the Public and the Administration, no individual decision shall be made on the exclusive basis of this algorithm". Conseil Constitutionnel, Judgement of 12 June 2018, No. 2018-765 DC, paragraph 70.

¹⁴⁸ Code of Public Procurement, Article 30(2)(a). See G. AVANZINI, *Article 30*, in R. VILLATA, M. RAMAJOLI, *Commentario al codice dei contratti pubblici. D.Lgs. 31 marzo 2023*, n. 36.

¹⁴⁹ C. COGLIANESE, *A framework for governmental use of machine learning*, Report to the Admin. Conf. of the U.S., December 2020. Available at: <https://www.acus.gov/sites/default/files/documents/Coglianese%20ACUS%20Final%20Report%20w%20Cover%20Page.pdf>.

¹⁵⁰ Italian Council of State, Section IV, 4 June 2025, No. 4857.

¹⁵¹ *Ibidem*.

administrative activity may no longer be true in relation to more sophisticated ones. Indeed, extensive research has shown that advanced forms of AI possess the potential to influence, and in some cases even supplant, human judgment, a phenomenon commonly identified as automation bias¹⁵². Second, following this line of reasoning, Article 30 of the Code of Public Procurement appears inapplicable, in its entirety, in cases where an AI system is employed to generate predictions or recommendations based on which the public authority takes subsequent steps or makes decisions¹⁵³.

As highlighted throughout this research, in the European context, law enforcement authorities – as well as public administrations more broadly – employ AI systems primarily as supportive tools for subsequent actions carried out by human officers, rather than as autonomous decision-makers. This is precisely the case with predictive policing, where law enforcement authorities employ predictive software to extract patterns from data, expand their knowledge base – often through the elaboration of risk assessments – and subsequently make decisions that may involve heightened surveillance of specific individuals or locations¹⁵⁴. In these cases, therefore, AI systems do not make the final decision – the one that directly affects the individual – but instead shape and orient subsequent human-led monitoring and decision-making processes.

The reading outlined above implies, by contrast, that in cases where AI systems do not directly determine outcomes, no strict obligation arises to guarantee the availability of the source code, its documentation, or other elements necessary for the interpretation of the system’s functioning. It is this author’s submission, however, that such a conclusion cannot be endorsed in its entirety. In particular, a distinction must be drawn between the availability of the source code to public administrations and applicants’ access to it. Indeed, public administrations must have access to the source code and must therefore guarantee its availability during the procurement process. In this respect, Article 30(2)(a) should be construed as a provision of general application, obliging administrations that procure AI systems to secure such availability¹⁵⁵. By contrast, applicants’ access

¹⁵² K. GODDARD ET AL., *Automation bias: a systemic overview of frequency, effect mediators, and mitigators*, in *Journal of the American Medical Informatics Association*, 2012, 121; L. J. SKITKA, K.L. MOSIER, M. BURDICK, *Does automation bias decision making?*, in *International Journal of Human-Computer Studies*, 1999, 991. It is also referred to as “effect moutonnier”. A. GARAPON, J. LASSÈGUE, *Justice Digitale. Révolution graphique et rupture anthropologique*, Presses Universitaires de France, 2018.

¹⁵³ Article 30, paragraph 2, letter (a), codifies the principle established in the most recent case law of the Council of State concerning the adoption of so-called “algorithmic decisions.” In such cases, the source code should be regarded as an administrative act or an administrative document and, as such, is subject to the right of access pursuant to Law No. 241 of 1990. The possible existence of intellectual property rights over the software, held by a private contractor, cannot therefore constitute an obstacle to such access. Italian Council of State, Judgement No. 4857/2025. By contrast, Article 30(3) correctly applies only to algorithmic decisions.

¹⁵⁴ Similar considerations apply to algorithmic video surveillance tools, which are employed to detect or identify individuals; however, subsequent actions are, at least for the time being, carried out by human officials rather than by the AI system itself.

¹⁵⁵ AVANZINI, *Article 30*, cited, 150. Moreover, Article 30(2) makes no reference to “decisions taken through automation,” unlike Article 30(3), which explicitly does so.

to the source code is a different matter. As clarified in the case law, this access is limited to instances where decisions are made by the software without human intervention¹⁵⁶. Indeed, when decisions are supported – but not determined – by AI systems, the affected individual may be less concerned with the source code than with how the system’s output was integrated into the socio-technical decision-making process and how it influenced subsequent human actions.

In any case, the cited jurisprudence and legislation have clearly expressed a preference for open-source solutions over proprietary software. More broadly, this orientation is consistent with the growing trend in EU policies and regulations, as well as in many national frameworks, to promote open-source adoption¹⁵⁷. The academic and policy literature on the use of open-source software in public administrations is by now extensive¹⁵⁸. It is thus already well established that the tendency in the European landscape is to ensure the disclosure of the source code as a way to achieve transparency¹⁵⁹. However, the issue of source code disclosure remains particularly sensitive. On the one hand, requiring disclosure effectively overrides intellectual property protections when private entities collaborate with public administrations, potentially discouraging some companies from participating. On the other hand, full transparency may enable adversaries or malicious actors to game the system. This risk is considered particularly acute in systems designed to detect patterns of behaviour, which may be deliberately manipulated through sophisticated adversarial attacks intended

¹⁵⁶ There must be a nexus of instrumentality between the right of access and the “final” legal situation, meaning that the disclosure of the administrative document must, on the basis of an *ex ante* prognostic assessment, be regarded as a means – instrumental in this sense – for acquiring evidentiary elements concerning the facts (both principal and secondary) that constitute the disputed “final” legal situation and the corresponding claims that may be asserted in court. The appellant company reached this conclusion on the premise that the algorithm used by *Digitalpa S.r.l.* constitutes an IT system substituting an ordinary procedural sequence and, as such, is strictly functional to the operative content of the decision by which the appellant was excluded from the tender.

¹⁵⁷ At the European level, see European Commission, *Strategy on the Software Open Source 2020-2023*, 21 October 2020, COM(2020) 7149 final. At the national level, the Italian Agency for the Digital (AgID) has issued guidelines which oblige public administrations and providers of IT solutions and programmes to make their source code available, in order to encourage the reuse of a software solution, i.e. its use in a context other than that for which it was originally developed. AgID, *Guidelines for the acquisition and reuse of software for public administrations*, February 2020. Available at: <https://docs.italia.it/italia/developers-italia/gl-acquisition-and-reuse-software-for-pa-docs/en/stabile/index.html#>.

¹⁵⁸ For an interesting discussion on open source software in public administrations see S. DEL GATTO, *I sistemi proprietari, l’open source e la pubblica amministrazione*, in *Giornale di diritto amministrativo*, 5, 2021, 571. The preference for open-source solutions is in line with Article 68, paragraph 1-ter) CAD, which allows public administrations to resort to proprietary software only when free and open-source alternatives are unavailable. A.G. OROFINO, *Open source e pubblica amministrazione*, in G. CASSANO (Ed.), *Diritto delle nuove tecnologie informatiche e dell’Internet*, Ipsoa, 2002, 1317; F. FRACCHIA, *Open source e pubblica amministrazione*, in M. BERTANI (Ed.), *Open source*, Giuffrè, 2005, 216; F. BRAVO, *Software “open source” per la p.a. tra diritto d’autore, appalti pubblici e diritto dei contratti. La licenza pubblica dell’UE per i programmi a codice sorgente aperto*, in *Diritto dell’Informazione e dell’Informatica*, 2008, 865. From a different perspective, in the Dutch administration, see A. VAN LOON, D. TOSHKOV, *Adopting open source software in public administration: The importance of boundary spanners and political commitment*, in *Government Information Quarterly*, 2015, 32(2), 207-215.

¹⁵⁹ C. CHAUSSON, *France opens the source code of tax and benefits calculators to increase transparency*. European Commission, November 2016. Available at: <https://joinup.ec.europa.eu/collection/egovernment/document/france-opens-source-code-tax-and-benefits-calculators-increase-transparency>.

to mislead AI models¹⁶⁰. Equally critical are the implications concerning data security¹⁶¹. For these reasons, it remains questionable whether the preference for open-source solutions can be appropriately extended to law enforcement contexts¹⁶². Notwithstanding these concerns, a report published by EUROPOL stated that promoting open-source AI systems could foster a community-driven approach to development, thereby enhancing both the quality and applicability of such tools. Furthermore, it may broaden access to advanced technologies, enabling even smaller law enforcement entities to harness the potential of AI¹⁶³.

Unlike Europe, United States adopts a distinctive approach: rather than mandating full disclosure of the source code, they require the disclosure only of those elements of the software that are deemed relevant in light of the specific interests at stake¹⁶⁴. Indeed, the mere availability of the source code reflects only a narrow facet of transparency and risks obscuring a more comprehensive understanding of what transparency entails¹⁶⁵. A genuinely comprehensive conception of transparency must extend to the overall design of the system, including its intended purposes, the methodological and normative choices underlying data selection, the criteria informing model architecture and parameterisation, and, more generally, the positioning of the AI system within the socio-technical process in which it operates.

In any case, it remains, in my assessment, incontrovertible that public authorities – and law enforcement authorities specifically – must be vested with access to the source code, together with comprehensive information on the system’s architecture and functioning. In particular, such access ought to be safeguarded through the stipulation of appropriate procurement requirements and contractual clauses¹⁶⁶. In this manner, algorithmic systems that fail to satisfy predetermined standards of transparency are precluded from acquisition within the public sector, while non-compliance by

¹⁶⁰ Such considerations have been highlighted in Ada Lovelace Institute, AI Now Institute and Open Government Partnership, *Algorithmic accountability for the public sector*, cited, 46.

¹⁶¹ Data security issues become relevant in relation to the GDPR.

¹⁶² The use of open-source solutions in law enforcement shall not be confounded with open-source intelligence. CEPOL, 33/2025/ONS: *Open-Source Intelligence (OSINT) and IT Solutions - Train the Trainer*, 2025. Available at: <https://www.cepola.europa.eu/training-education/33-2025-ons-open-source-intelligence-osint-and-it-solutions-train-trainer>.

¹⁶³ EUROPOL, *AI and policing. The benefits and challenges of artificial intelligence for law enforcement*, cited.

¹⁶⁴ C. COGLIANESE, E. LAMPMANN, *Contracting for algorithmic accountability*, Administrative Law Review Accord, 6(2), 2021, 175-206. Available at: https://scholarship.law.upenn.edu/faculty_scholarship/2311/.

¹⁶⁵ Ada Lovelace Institute, AI Now Institute and Open Government Partnership, *Algorithmic accountability for the public sector*, cited, 47.

¹⁶⁶ As recognised by Ada Lovelace Institute, AI Now Institute and Open Government Partnership, *Algorithmic accountability for the public sector*, cited, 33, “Government procurement conditions have been an important area of intervention for transparency and accountability”. Among other works, see Open Government Partnership, *Open Contracting and Public Procurement*. Available at: <https://www.opengovpartnership.org/policy-area/open-contracting/>. World Economic Forum, *AI Government Procurement Guidelines*, 2020. Available at: <https://www.weforum.org/reports/ai-procurement-in-a-box/ai-government-procurement-guidelines>. UK Government, *Guidelines for AI Procurement*, 2020. Available at: <https://www.gov.uk/government/publications/guidelines-for-ai-procurement>. municipality Amsterdam, *Standard Clauses for Municipalities for Fair Use of Algorithmic Systems*, 2020. Available at: <https://www.amsterdam.nl/innovatie/>.

vendors with contractual obligations gives rise to contractual liability. Moreover, procurement instruments may be framed so as to enable interventions both at the design stage of algorithmic systems and throughout their subsequent operational lifecycle. The incorporation of such clauses into tender documents, however, presupposes that public authorities – and, with particular regard to the present analysis, law enforcement authorities – possess the requisite technical and legal expertise to negotiate effectively with private vendors¹⁶⁷. In any case, a principal advantage of in-house system development resides in the capacity to design AI systems in alignment with the public administration’s specific operational and normative requirements, while embedding public value considerations, including transparency imperatives. This approach mitigates the risk of value misalignment inherent in AI systems developed by private entities whose objectives are predominantly oriented toward profit maximization¹⁶⁸. Nonetheless, such solutions are not invariably feasible, given constraints relating to budgetary limitations, personnel resources, or requisite technical expertise.

3.3. Human oversight

Ensuring transparency assumes particular significance in relation to the human oversight requirement introduced for high-risk AI systems in Article 14¹⁶⁹. Human oversight was initially established as an ethical principle governing the responsible development and deployment of AI systems¹⁷⁰. From a

¹⁶⁷ The reference inevitably concerns the obligation to promote AI literacy. The French Council of State, in its report *Intelligence artificielle et action publique: construire la confiance, servir la performance*, acknowledged that “l’externalisation requiert également des compétences internes afin de définir le cahier des charges, de sélectionner et de piloter les prestataires. S’agissant de la sélection, il est indispensable que l’administration comprenne ce qu’elle achète exactement ». Ada Lovelace Institute, *Buying AI*, cited.

¹⁶⁸ C. CASONATO, *L’intelligenza artificiale fra pubblico e private: una sfida per il costituzionalismo*, in *DPCE*, 1, 2025, 5-13; MARCHETTI, TORCHIA, *AI and Public Administration*, cited.

¹⁶⁹ For a discussion on the human oversight requirement see G. LAZCOZ MORATINOS, *Human Oversight Or Monitoring In Article 14 Of The Artificial Intelligence Act: A Mere Mandatory Requirement For High-Risk Systems?*, L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 717-737.

¹⁷⁰ From an ethical perspective, the legal requirement of ‘human oversight’ can be framed in three forms: human-in-the-loop (HITL, human intervention), human-on-the-loop (HOTL, human supervision), and human-in-command (HIC, human control) In particular, the HLEG specified that the HITL approach should allow for human intervention in every decision-making cycle of the system, the HOTL approach would involve human intervention during the system design cycle and monitoring of its operation, and the HIC approach would involve control of the AI system’s activity as a whole, including its overall economic, social, legal and ethical effects, as well as the ability to decide when and how to use the system in particular situations. A critical reading of Article 14 suggests, however, that the European legislator may have excluded HITL, privileging instead HOTL and HIC. In fact, while Article 14 and Article 26(2) require human intervention at both the design stage and during the use of an AI system, they devote insufficient attention to the decisional phase delegated to the system itself, precisely the domain in which HITL operates. Yet HITL, understood as continuous interaction between human beings and the system throughout its lifecycle, appears indispensable, offering a far stronger safeguard than HOTL or HIC, particularly in sensitive domains. At the same time, the full implementation of HITL is objectively difficult to guarantee in contexts involving machine learning or deep learning, given that these technologies, by their very nature, render continuous human control considerably more complex. The HLEG itself acknowledged in its *Ethical Guidelines* that achieving such a goal is in many cases neither possible nor desirable. HLEG, *Ethics Guidelines for Trustworthy AI*, cited, 16.

regulatory perspective, the precursor to the human oversight requirement is found in the “principle of non-exclusivity” and in the right not to be subject to decisions based solely on automated processing where such decisions produce legal effects concerning an individual or significantly affect them, as enshrined in Articles 22 of the GDPR and 11 of the LED¹⁷¹. Accordingly, data protection legislation requires that some form of human intervention be ensured in order for individual automated decision-making to be considered lawful within the European territory¹⁷². In a comparable manner, the AI Act provides that high-risk AI systems shall be designed and developed so as to guarantee effective human oversight across their entire lifecycle¹⁷³. This entails that a human operator must at all times be in place to perform oversight functions, endowed with both the competence and the authority necessary to exercise effective control over the system. This marks a fundamental difference between data protection legislation and the AI Act. Under the former – namely Articles 22 GDPR and 11 LED – the safeguards apply when a decision is taken about an individual on the basis of an automated processing. Article 14 of the AI Act, by contrast, operates on a broader functional trigger: it applies whenever a high-risk AI system is used to support a decision-making process, irrespective of whether the resulting decision is ultimately automated. However, the scope of Article 14 is simultaneously narrower in another respect, as it applies exclusively to high-risk AI systems. Articles 22 GDPR and 11 LED, on the other hand, impose their respective safeguards more generally, without being limited to a predefined risk category.

¹⁷¹ Building on this principle established in data protection legislation, the Italian Council of State recognised it as a safeguard for citizens against algorithmic decision-making. See A. SIMONCINI, *Amministrazione digitale algoritmica. Il Quadro costituzionale*, in R. CAVALLO PERIN, D.-U. GALETTA, *Il Diritto Dell'amministrazione Pubblica Digitale*, Giappichelli, 2020. A. ROIG, *Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR)*, in *European Journal of Law and Technology*, 8(3), 2017, 1-17.

¹⁷² The Article 29 Working Party (WP29, now EDPB) has underscored in its Guidelines that, in order to avoid triggering the application of Article 22, human oversight must be *genuine and substantive*. This requires that the overseer not only possess the requisite authority but also the necessary competence to effectively review, and where appropriate, modify or overturn the decision. WP29, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, 3 October 2017. In practice, Article 22 has seen limited application. The limited practical application of the provision is primarily attributable to the fact that the principle of non-exclusivity is subject to numerous exceptions. The circumstances in which the prohibition does not apply are set out in the provision itself, which prescribes in such cases additional safeguards aimed at protecting the rights, freedoms, and legitimate interests of the data subject. Among the safeguards referred to is “*the right to obtain human intervention on the part of the controller, to express his or her point of view, and to contest the decision*”. The breadth of the exceptions to which it is subject has, in fact, rendered the protective purpose of the rule virtually ineffective. Cfr. A. SIMONCINI, *Verso la regolamentazione della Intelligenza Artificiale. Dimensioni e governo*, in *BioLaw Journal*, 2, 2021, 416.

¹⁷³ See B. MARCHETTI, *La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica*, in *BioLaw Journal*, 2, 2021, 377. It is to be hoped that Article 14 of the AI Act will find broader application than Article 22 GDPR which, as noted, has remained largely a “dead letter”. Nevertheless, the provision appears promising, as it does not include exceptions, unlike the GDPR rule, where such exceptions represented the main reason for its limited application. Several commentators have noted that the drafting of Article 14 raises critical concerns that warrant closer scrutiny.

The problem of deploying AI systems in the decision-making process – whether as a support tool for human decision-making, or, more critically, as a complete substitute for it – is by now well known¹⁷⁴. The risk lies in the considerable practical influence exerted by AI systems¹⁷⁵, which arises from several factors: the human inability to fully comprehend their functioning and thus to make truly informed decisions about them; their perceived scientific or neutral character¹⁷⁶; and the practical convenience of aligning one’s decision with the system’s output rather than departing from it¹⁷⁷. The result is a *de facto* delegation of responsibility, since decision-makers are unlikely to depart from the system’s output¹⁷⁸.

Indeed, scholars have extensively discussed the risk of the “machine” capturing human judgment and decision-making¹⁷⁹. Aware of this phenomenon, the European legislator explicitly states in Article 14 of the AI Act that natural persons entrusted with oversight functions must remain cognizant of the risk of automation bias to which they may be exposed¹⁸⁰. This phenomenon in human-machine interaction describes the tendency of individuals to exhibit unwarranted reliance – showing the tendency to conform, sometimes “automatically” or “excessively” – on the output of AI systems, thereby diminishing critical scrutiny and increasing the risk of overlooking potential errors¹⁸¹.

However, on this point, the AI Act could arguably have adopted a more assertive stance. Notably, the provision avoids employing terms such as “aid”, “support” or “instrumental” in reference to AI systems, thus rendering ambiguous whether these systems are envisioned as ancillary mechanisms or as potential surrogates for human judgment¹⁸². While this acknowledgment may not be decisive in

¹⁷⁴ A. SIMONCINI, *Verso la regolamentazione della Intelligenza Artificiale*, cited, 416 e 417; Cfr. MARCHETTI, *La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica*, cited.

¹⁷⁵ Once an automated decision-making system is introduced into a human decision-making process, the automated system tends, over time, to take over the entire decision itself. In these terms, SIMONCINI, *Verso la regolamentazione della Intelligenza Artificiale*, cited, 416.

¹⁷⁶ *Ibidem*. Particularly noteworthy in this regard is MARCHETTI, *La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica*, cited, 373 which underscores the machine’s tendency to exert a significant influence on the human decision-maker – an anchoring effect – leading the latter to adjust their judgment when it diverges from that produced by an AI system.

¹⁷⁷ C. CASONATO, *Per un’intelligenza artificiale costituzionalmente orientata*, in A. D’ALOIA (Ed.), *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, Franco Angeli, 2020, 104.

¹⁷⁸ As Maria Zanichelli has emphasized, however, the very act of deciding should constitute one of the main dividing lines for determining what human intelligence can delegate to artificial intelligence and what it cannot. M. ZANICHELLI, *Ecosistemi, opacità, autonomia: le sfide dell’intelligenza artificiale in alcune proposte recenti della Commissione europea*, in D’ALOIA (Ed.), *Intelligenza artificiale e diritto*, cited, 79. The author observes that the issue is also linguistic in nature. By ascribing to intelligent artifacts terms such as “decision”, “choice” or “action” and verbs such as “to act”, “to choose” and “to decide” we risk normalizing – albeit in a clearly improper manner – the perception that machines are capable of initiative and agency. This lexical slippage subtly shapes our conceptual framework, fostering an acceptance of the delegation of functions that are, in essence, constitutive of human autonomy.

¹⁷⁹ MARCHETTI, *Amministrazione digitale*, cited, 98. A GARAPON, J. LESSÈGUE, *Justice digitale*, cited, 239.

¹⁸⁰ AI Act, Article 14(4)(b).

¹⁸¹ An interesting example of how human decision-makers rely on AI system’s outputs is shown in relation to COMPAS. See M. VACCARO, J. WALDO, *The effects of mixing machine learning and human judgment*, in *Communication of the ACM*, 2019, 62, 11, 104–110. Available at: <https://doi.org/10.1145/3359338>.

¹⁸² In the justice sector, for instance, the importance of preserving the role of the judge within the adjudicatory process, and of safeguarding the judge’s constitutional guarantee function, has been emphasized by A. D’ALOIA, *Il diritto verso*

itself, the provision should be interpreted not only in its entirety but also in conjunction with earlier provisions, particularly those related to the principle of transparency: if the human overseer is able to comprehend the rationale or mechanisms by which the AI system arrives at its decisions, effective oversight becomes feasible, thereby reinforcing the human-in-the-loop safeguard. For example, consider a person-based AI system used in a law-enforcement context to assist officers in assessing the potential risk associated with an individual during an operational decision, such as determining whether additional supervisory review is required. To exercise meaningful oversight, the officer must be able to understand, at a minimum, the general categories of information the system relies upon – such as past recorded interactions, contextual behavioural indicators, or relevant situational factors – and how these contribute to the system’s risk assessment. With that level of insight, the officer can critically evaluate whether the system’s output is plausible, identify potential inaccuracies or unwarranted inferences, and ultimately decide whether to rely on, adjust, or disregard the suggestion. From a by design perspective, high-risk AI systems should be designed and developed to ensure that natural persons can effectively oversee their operation throughout the period of use, thereby enabling meaningful human supervision. To that end, the provider must identify adequate human oversight measures and integrate them into the system before it is placed on the market – such as providing the high-risk AI system with interfaces tools allowing the human operator to oversee the AI system – or capable of being implemented subsequently by the user¹⁸³. To this end, the instructions for use discussed above must include also the human oversight measures put in place by the provider¹⁸⁴.

Although the AI Act places primary responsibility on providers to design systems enabling effective human oversight, it likewise establishes complementary obligations for deployers. Deployers shall use the AI system in accordance with the instructions for use accompanying the system regarding human supervision measures. Throughout the entire operational phase of the system, the natural person entrusted with oversight must be able to: i) monitor the system throughout its life cycle; ii) identify malfunctions and anomalies in order to intervene promptly; iii) correctly interpret the system’s outputs; iv) determine the circumstances under which the system should not be used; v) disregard, override or reverse the system’s outputs; and, vi) where necessary, intervene in or even interrupt its functioning. Human oversight, therefore, must be understood as encompassing the capacity to decide whether to employ an AI system in a given context, to determine the extent of

“il mondo nuovo”. *Le sfide dell’Intelligenza Artificiale*, in D’ALOIA (Ed.), *Intelligenza artificiale e diritto*, cited, 37. Likewise, Carlo Casonato has highlighted the unsuitability of entirely replacing human decision-making with artificial outputs, arguing for a constitutionally oriented approach to AI. CASONATO, *Per un’intelligenza artificiale costituzionalmente orientata*, cited.

¹⁸³ Information about such measures shall be given to the deployer in the instructions for use.

¹⁸⁴ AI Act, Article 13(3)(d).

human discretion during its operation, and to override a decision produced by the system itself¹⁸⁵. Recognising the potential adverse impacts of high-risk AI systems listed in Annex III point 1(a), namely biometric identification, the AI Act requires that oversight mechanisms prevent deployers from acting on any identification result unless it has been separately verified and confirmed by at least two qualified natural persons with appropriate training and authority. Nonetheless, this double-verification requirement may be waived for AI systems used in law enforcement, migration, border management, or asylum contexts, where Union or national law considers its application disproportionate¹⁸⁶.

In order to ensure the effectiveness of such obligations, according to Article 26(2) of the AI Act, human oversight must be entrusted to natural persons who possess the requisite competence, training, and authority to intervene effectively in the operation of the AI system. Moreover, these individuals must be provided with the necessary organisational support and resources to ensure that oversight is not merely formal, but genuinely capable of preventing or mitigating risks to health, safety, and fundamental rights¹⁸⁷. In practice, the above obligations entail that the natural person to which oversight is entrusted must be able to interpret outputs critically, challenge system recommendations, and override them when appropriate. To this end, the obligation of “AI literacy”¹⁸⁸, directed towards the diffusion of technical expertise among the provider or deployer’s staff and other persons dealing with the operation and use of AI systems on their behalf, is crucial¹⁸⁹. Proper training and preparation

¹⁸⁵ HLEG, *Ethics Guidelines on Trustworthy AI*, cited, 16. Article 14 of the AI Act highlights that the natural person to whom human oversight is entrusted shall decide, in any particular situation, not to use the high-risk AI system or to otherwise disregard, override or reverse the output of the high-risk AI system (letter d) and to intervene in the operation of the high-risk AI system or interrupt the system through a ‘stop’ button or a similar procedure that allows the system to come to a halt in a safe state (letter e).

¹⁸⁶ AI Act, Article 14(5); Recital 73: “*This requirement should not pose unnecessary burden or delays and it could be sufficient that the separate verifications by the different persons are automatically recorded in the logs generated by the system. Given the specificities of the areas of law enforcement, migration, border control and asylum, this requirement should not apply where Union or national law considers the application of that requirement to be disproportionate*”.

¹⁸⁷ It should be made clear that neither humans nor human decision-making ought to be idealized. A.Z. HUQ, *A Right to a Human Decision*, in *Virginia Law Review*, 106(3), 2020, 611-688.

¹⁸⁸ The concept of AI literacy mentioned in article 4 of the AI Act relies on the definition of the term given in article 3(56) of the AI Act, according to which: “*‘AI literacy’ means skills, knowledge and understanding that allow providers, deployers and affected persons, taking into account their respective rights and obligations in the context of this Regulation, to make an informed deployment of AI systems, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause*”. The European Artificial Intelligence Board should “*support the Commission, to promote AI literacy tools, public awareness and understanding of the benefits, risks, safeguards, rights and obligations*”. AI Act, Recital 20, See *Living repository to foster learning and exchange on AI literacy*. Available at: <https://digital-strategy.ec.europa.eu/en/library/living-repository-foster-learning-and-exchange-ai-literacy>. AI literacy is a general obligation applicable to both private entities and public authorities in relation to all AI systems, regardless of their risk level.

¹⁸⁹ The link of AI literacy with transparency and human oversight has been highlighted by the European Commission. See the cited *AI Literacy - Questions & Answers*. Available at: <https://digital-strategy.ec.europa.eu/en/faqs/ai-literacy-questions-answers>.

of staff is indeed a prerequisite for governing AI systems and assessing their implications in their intended context¹⁹⁰.

Moreover, it is crucial to establish the appropriate scope of delegation to AI systems, as well as to define the decision-making areas that must necessarily remain within the remit of human agents. Luciano Floridi underscores the need to distinguish between the “delegation of processes” and the “delegation of decisions”¹⁹¹. Whereas the delegation of processes, provided adequate safeguards are in place, does not in principle give rise to excessive difficulties, the delegation of decisions must be approached with the utmost caution. In fact, the preservation of the ultimate authority to confirm, reject, or amend a decision generated by or with the support of an AI system constitutes not only a human right but also a duty, insofar as it ensures the retention of responsibility for one’s actions¹⁹². Indeed, accountability can be ensured only insofar as the ultimate decision-making authority remains vested in the human decision-maker¹⁹³. The concept of algorithmic accountability has become central to the discourse on AI in the public sector¹⁹⁴, as it seeks to ensure that those who procure, and deploy AI systems are held responsible for, and ultimately answer to, their potential negative impacts. More precisely, this notion of accountability entails not only that actors are capable of justifying and are answerable for their use of algorithmic systems, but also that they may face consequences arising from such use¹⁹⁵. In a noteworthy study conducted by EUROPOL, accountability is identified as one of the most important values for an ethical and legitimate use of AI systems in law enforcement. Indeed, in relation to the use of AI systems accountability is described as “*a proxy measure for the extent of their legitimacy within [...] communities*”¹⁹⁶. Clearly, it is transparency that enables the

¹⁹⁰ For a comparative perspective on the training of civil servants in relation to AI see DECAROLIS, MARCHETTI, TORCHIA (Eds.), *The EU Digital Regulation and its Impact on Member States*, cited.

¹⁹¹ P. FABBRI, *AI ed etica: relazione delicata, ma cruciale. Intervista a Luciano Floridi*, 12 February 2019. Available at: <https://www.patriziafabbri.it/wp/2019/02/ai-ed-etica-relazione-delicata-ma-cruciale-intervista-a-luciano-floridi/>.

¹⁹² Cfr. M. ZANICHELLI, *Ecosistemi, opacità, autonomia*, cited, 85.

¹⁹³ Accountability is a principle recognised in many legal and policy documents. *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, Article 9. F. DOSHI-VELEZ, ET AL., *Accountability of AI under the law: The role of explanation*, Berkman Klein Center for Internet & Society working paper, 2017. “By exposing the logic behind a decision, explanation can be used to prevent errors and increase trust. Explanations can also be used to ascertain whether certain criteria were used appropriately or inappropriately in case of a dispute”. *Ibidem*, 2. T.M. LECHTERMAN, *The Concept of Accountability in AI Ethics and Governance*, in J. B. BULLOCK ET AL. (Eds.), *The Oxford Handbook of AI Governance*, Oxford, 2022.

¹⁹⁴ For a more general discussion on algorithmic accountability, see Ada Lovelace Institute et al., *Algorithmic accountability for the public sector*, cited. For a discussion of the principle of accountability in the specific context of law enforcement see EUROPOL, *Accountability Principles for Artificial Intelligence (AP4AI) in the Internal Security Domain*, February 2022. Available at: https://www.europol.europa.eu/cms/sites/default/files/documents/Accountability_Principles_for_Artificial_Intelligence_AP4AI_in_the_Internal_Security_Domain.pdf.

¹⁹⁵ MARCHETTI, *Amministrazione digitale*, cited, 105.

¹⁹⁶ EUROPOL, *Accountability Principles for Artificial Intelligence (AP4AI) in the Internal Security Domain*, cited, 14. The importance of establishing police accountability has been highlighted also in other documents. UNODC, *Handbook on Police Accountability, Oversight and Integrity*, United Nations, July 2011; I. KEARNS, R. MUIR, *Data-driven Policing and Public Value*, March 2019. Available at: https://www.police-foundation.org.uk/wp-content/uploads/2010/10/data_driven_policing_final.pdf.

affected individual or the public to use information to hold public agencies accountable for their deployment of algorithmic systems¹⁹⁷.

As a concluding remark on the topic, the close connection between human oversight and accountability inevitably raises questions about the limitation of the former to high-risk AI systems, while for other systems its application is merely encouraged through voluntary Codes of Conduct. In this author's contention, however, considering the responsibilities and principles that govern administrative action, such a requirement should reasonably apply to all AI systems employed by public authorities, regardless of their designated level of risk¹⁹⁸.

3.4. The tactical barriers and the difficulty in ensuring the right to an effective remedy

Within the sphere of law enforcement, the aspiration to full transparency is constrained by the so-called "*tactical barriers*"¹⁹⁹. Excessive disclosure regarding the design, deployment, or functioning of AI systems used for law enforcement purposes may compromise investigative strategies, facilitate circumvention by targeted individuals, or otherwise diminish the operational value of these tools. In recognition of this tension, Article 49(4) of the AI Act establishes that for high-risk AI systems in law enforcement – as well as migration, asylum and border control management – the registration referred to in that provision shall be in a secure non-public section of the EU database²⁰⁰. In addition, the registration shall include only the information identified in Article 49(4), as discussed above.

Although the provision ostensibly seeks to balance transparency, conceived in terms of information obligations, with the imperative of protecting the confidentiality and integrity of law enforcement activities, in practice, the balance overwhelmingly favours confidentiality to the detriment of meaningful transparency. In particular, although limitations on the information to be disclosed are frequently justified on operational or tactical grounds inherent to the functioning of law enforcement activities, the requirement that such (already limited) information be registered in a secure, non-public section of the EU database – with access restricted solely to the Commission and competent national authorities – remains open to debate. Such restrictions effectively exclude the most sensitive uses of AI systems from full public scrutiny. In any case, the call for greater transparency in the use of technological tools by law enforcement authorities persists²⁰¹, being considered a means to ensure

¹⁹⁷ This may be possible only in the case where the audience is able to understand the information and effectively use them. J. KEMPER, D. KOLKMAN, *Transparent to whom? No algorithmic accountability without a critical audience*, in *Information, Communication & Society*, 2019, 22(14), 2081-2096.

¹⁹⁸ In these terms, see MARCHETTI, TORCHIA, *AI and Public Administration*, cited.

¹⁹⁹ A.G. FERGUSON, *Illuminating Black Data Policing*, cited, 14.

²⁰⁰ AI Act, Article 71.

²⁰¹ E. BAKKE, *Predictive policing: the argument for public transparency*, in *NYU Annual Survey of American Law*, 2018, 74, 147. S.D. ZANSBERG, P. CAMPOS, *Sunshine on the Thin Blue Line: Public Access to Police Internal Affairs Files*, Communications Lawyer, 2004, 34.

procedural fairness²⁰², accountability²⁰³ and eventually broadening trust in law enforcement operations²⁰⁴.

The fact that the European legislator has opted for particularly restrictive provisions on institutional transparency does not preclude national legislators from imposing more extensive transparency requirements, for instance, by obliging law enforcement authorities to publish certain information regarding the AI systems they employ, albeit without disclosing sensitive operational details²⁰⁵. It therefore appears necessary to strike an appropriate balance between the information that ought to be made publicly available and the confidentiality requirements arising from competing interests, thereby avoiding full disclosure, which may prove counterproductive, particularly in terms of maintaining public trust²⁰⁶. During the deliberations on the New York City Police Offender Surveillance Technology (POST) Act²⁰⁷ experts testified before the City Council that, while the public disclosures mandated by the Act would enhance public insight into the surveillance technology used by the NYPD, they remain insufficiently detailed to enable adversaries to exploit the systems or otherwise endanger public safety²⁰⁸. Moreover, in areas where transparency is difficult to achieve due to compelling interests, such as crime prevention and detection, the establishment of independent expert committees to monitor deployed AI systems has been suggested as a valuable way to balance oversight with the confidentiality of sensitive operational data²⁰⁹.

²⁰² Transparency help identifying errors and biased results, allowing better oversight and thus fairer results.

²⁰³ The externalization of responsibilities represents one of the biggest problems in the reliance of law enforcement authorities on private providers. *“The information disparity between law enforcement and developers results in an unchecked delegation of democratically granted authority to a private party. Public transparency resolves this issue by providing a check on developers”*.

²⁰⁴ In the American literature, a correlation between faith in the legal system and obedience to the law has been highlighted by T.R. TYLER, *Why People Obey The Law*, 1990, 4–5. T.R. TYLER, J. FAGAN, *Legitimacy and Cooperation: Why Do People Help the Police Fight Crime in Their Communities?*, in *Ohio State Journal of Criminal Law*, 6, 2008, 231–275. Therefore, if people believe that the legal system – and thus the law – is legitimate, they tend to comply with it.

²⁰⁵ It must be remembered that the AI Act only sets out the minimum requirements for the admissibility of AI systems within the European legal order. MARCHETTI, TORCHIA, *AI and Public Administrations*, cited. Law No. 132/2025, adopted in the Italian context, delegates to the government the task of *“the establishment of specific regulations governing the use of artificial intelligence systems in policing activities”*. AI Act, Article 24(2)(h).

²⁰⁶ Ada Lovelace Institute et al., *Algorithmic accountability for the public sector*, cited, 46.

²⁰⁷ An Assessment of NYPD’s Response, November 2022. Available at: https://www.nyc.gov/assets/doi/reports/pdf/2022/POSTActReport_Final_11032022.pdf. T. KAMRAN, *Testimony: NYPD’s Implementation of the POST Act*, 21 February 2025. Available at: <https://bds.org/latest/testimony-nypds-implementation-of-the-post-act>.

²⁰⁸ Ada Lovelace Institute et al., *Algorithmic accountability for the public sector*, cited, 46. It has been argued that one practice potentially beneficial for law enforcement’s use of AI is the establishment of algorithmic registers, as these may enhance public trust and acceptance of such technologies. In these cases, however, particular care must be taken to ensure that the information made public does not include sensitive policing data, the disclosure of which could compromise law enforcement operations and expose AI systems to exploitation or evasion by malicious actors. EUROPOL, *Accountability Principles for AI (AP4AI) in the Internal Security Domain: AP4AI Framework Blueprint*, cited, 31.

²⁰⁹ See e.g., Council of Europe, Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108), *Guidelines on Artificial Intelligence and Data Protection*, 25 January 2019, T-PD(2019)01. <https://unesdoc.unesco.org/ark:/48223/pf0000377881>. *“AI developers, manufacturers and service providers are encouraged to set up and consult independent committees of experts from a range of fields, as well as engage with independent academic institutions, which can contribute to designing human rights-based and ethically and socially-oriented AI applications, and to detecting potential bias. Such committees may play an especially*

By contrast, the restrictions on the right to an explanation under Article 86, which apply to individuals directly affected by decisions based on the output of an AI system, appear less convincing, as operational and tactical justifications may no longer be present when a decision with adverse consequences for an individual is taken. Indeed, for affected individuals, receiving meaningful information²¹⁰, specifically, an explanation of the role of the AI system in the decision-making process and the main elements underlying the decision taken, constitutes a necessary precondition for the effective exercise of their fundamental rights, including the right to challenge decisions and to seek compensation²¹¹.

National legal systems already provide avenues of redress for natural and legal persons whose rights and freedoms are adversely affected²¹². However, the effective exercise of such remedies may become significantly more difficult when AI systems are involved, given their opacity, and the challenges they pose to establishing causation²¹³. Indeed, it has been observed that significant difficulties persist in identifying and, ultimately, proving potential breaches of the law – including provisions safeguarding fundamental rights – as well as in attributing liability and meeting the legal conditions necessary to claim compensation²¹⁴. It is widely acknowledged that challenging the outcomes of AI systems is complex due to inherent technical complexities of such systems, to the lack of accessible information, or the courts’ limited knowledge and expertise in providing effective remedies in such a new landscape.

important role in areas where transparency and stakeholder engagement can be more difficult due to competing interests and rights, such as in the fields of predictive justice, crime prevention and detection”. Ibidem, 2.

²¹⁰ Information provided must be *usable*, taking into account both the volume of data disclosed and the technical complexity of the language in which it is conveyed. Ada Lovelace Institute et al., *Algorithmic accountability for the public sector*, cited, 47. Recital 171 of the AI Act states that “*explanations must be clear and meaningful, providing a basis on which affected individuals can exercise their rights*”.

²¹¹ AI Act, Recital 171. In general terms, affected individuals are primarily interested in receiving an explanation of the decision taken on the basis of the AI system’s output, either to understand how to adjust their conduct in order to secure a more favourable outcome in the future, or to grasp the reasoning underlying the decision so as to effectively challenge or contest it. Transparency is crucial to consider the use of an AI system as legitimate, especially when used in the public sector. K. DE FINE LICHT, J. DE FINE LICHT, *Artificial intelligence, transparency, and public decision-making*, in *AI & Society*, 35, 2020, 917-926.

²¹² The right to challenge decisions taken affecting oneself is a fundamental right. See Article 47 CFREU on the right to an effective remedy and to a fair trial. M.E. KAMINSKI, J. M. URBAN, *The Right to Contest AI*, in *Columbia Law Review*, 121 2021, 1957. Available at: <https://scholar.law.colorado.edu/faculty-articles/1393>. For an interesting discussion on the right to an effective remedy in the digital age see G. DE GREGORIO, S. DEMKOVA, *The Constitutional Right to an Effective Remedy in the Digital Age- A Perspective from Europe*, in C. VAN OIRSOUW, J. DE POORTER, I. LEIJTEN, G. VAN DER SCHYFF, M. STREMLER, M. DE VISSER (Eds), *European Yearbook of Constitutional Law*, Springer, 2024, 223-254.

²¹³ Chapter IX, Section 4 provides only a right to lodge a complaint with a market surveillance authority.

²¹⁴ It is extremely complex to “*prove possible breaches of laws, including legal provisions that protect fundamental rights, attribute liability and meet the conditions to claim compensation*”. European Commission, *White Paper on Artificial Intelligence: a European approach to excellence and trust*, 19 February 2020, 14.

Moreover, given the challenges inherent in contesting law enforcement operations, the restriction of information made available to affected individuals further exacerbates the difficulty of effectively exercising such rights²¹⁵.

Consider, for instance, the use of a real-time RBI system, as discussed in Chapter 2: a law enforcement authority may rely on such a system and, on the basis of a positive match, proceed to arrest an individual identified as the probable offender²¹⁶; in such circumstances, the affected individual encounters substantial obstacles in contesting the decision, not only because of the strict time constraints inherent in arrest procedures and the authoritative and immediate nature of law enforcement determinations but also owing to the lack of necessary information concerning the system employed²¹⁷. It therefore appears to be an essential element of due process that, where defence counsel requests disclosure of the evidence underpinning an arrest decision, access to the system's outputs and other relevant information must be granted and not withheld on grounds of confidentiality. As argued above, restrictions under Article 86 – where invoked, both at the supranational and national level – should be narrowly circumscribed and applied with particular caution. Even more critical is the circumstance described above in which an individual, on the basis of a risk assessment generated by a predictive policing software, may become subject to enhanced surveillance measures without such measures necessarily resulting in an arrest or any other form of physical restriction²¹⁸. Unlike in arrest cases, the individuals often do not know they are under enhanced surveillance, making the exercise of rights largely theoretical unless a tangible consequence emerges (e.g., being stopped, interrogated, or refused access to a service). In this context, the exercise of judicial review proves even more difficult²¹⁹.

As previously noted, Article 86 presents multiple limitations, and similar concerns arise with respect to Article 26(11). In particular, the latter provision establishes an obligation for deployers of high-

²¹⁵ Analogous concerns arise in relation to the migration and security contexts. T. MOLNÁR, *Algorithmic Decision-Making in Migration and Security in the EU: Challenges in Ensuring Effective Legal Remedies*, in M. VARJU, K. MEZEI (Eds.) *The Challenges of Artificial Intelligence for Law in Europe*. CAILEIC 2022. Data Science, Machine Intelligence, and Law, Vol 6, Springer, Cham, 147-172.

²¹⁶ The first ruling on the use of a RBI system to identify and arrest a person is ECtHR, Judgement of 4 July 2023, Case *Glukhin v. Russia*, Application no. 11519/20. G. MOBILIO, *La Corte EDU condanna il ricorso alle tecnologie di riconoscimento facciale per reprimere il dissenso politico: osservazioni a partire dal caso Glukhin c. Russia*, in *DPCE Online*, 2024, 62(1), 695-705.. Available at: <https://doi.org/10.57660/dpceonline.2024.2107>.

²¹⁷ Of course, the individual is not left without safeguards. Criminal law introduces procedural safeguards, such as the right to be informed of the reasons for their arrest (Article 6 ECHR, Article 47–48 CFREU) and the right to the defence counsel to request the disclosure of evidence underpinning the decision.

²¹⁸ In such cases “While this may not lead to a direct shift of the burden of proof upon the suspect or accused individuals, it may still oblige them to have to prove the inaccuracy of the AI system's output, thus undermining the presumption of innocence as safeguarded by Directive (EU) 2016/343”. D. CASABURO, I. MARSH, *Ensuring fundamental rights compliance and trustworthiness of law enforcement AI systems: the ALIGNER Fundamental Rights Impact Assessment*, in *AI and Ethics*, 4, 2024, 1571.

²¹⁹ Affected individuals may not always be aware of the use of AI systems in the decision-making process and of the role of the system in the decision-making process. Moreover, there is limited availability of specialised legal support.

risk AI systems listed in Annex III – when such systems make or assist in making decisions concerning natural persons – to inform those individuals that they are subject to the use of a high-risk AI system, albeit within the limits set by Article 13 of the LED. Article 86 is intrinsically linked to Article 26(11), as the latter constitutes a necessary precondition for the effective exercise of the former²²⁰. The interpretative challenge, however, lies in determining whether the “*decision [...] taken [...] on the basis of the output*” of a high-risk AI system or the information that shall be given when a high-risk AI system is used to “*assist in making decisions*” should also encompass activities that do not strictly qualify as “decisions” in the traditional sense²²¹. This issue is particularly salient in contexts such as predictive policing, where the output may not culminate in an immediate or formal “decision”, yet nonetheless results in intensified surveillance of specific individuals, with potential implications for their fundamental rights.

The difficulty of sustaining such an interpretation is intimately connected to the question of what degree of reliance on the output of an AI system is required for a human decision to fall within the scope of application of Articles 26(11) and 86. While it is clear that these provisions are not confined to situations of fully automated decision-making, the degree of automation required for their application remains unsettled. Recital 171 offers some interpretative guidance in relation to Article 86, clarifying that the expression “*based on the output*” should be understood as “*based mainly upon the output*”²²². In any event, definitive clarification on this point will ultimately depend on the interpretative guidance issued by the European Commission and on the jurisprudence of the CJEU. The latter, in its ruling in *OQ v Land Hessen*²²³ affirmed that the notion of “decision” under Article 22 GDPR is to be interpreted broadly, extending to profiling outputs such as credit scores, when a third party draws strongly on that value to make decisions affecting the data subject. By extension, it is arguable that comparable considerations may be transposed to risk scores produced by predictive

²²⁰ In similar terms, M.E. KAMINSKI, G. MALGIERI, *The Right to Explanation in the AI Act*, U of Colorado Law Legal Studies Research Paper No. 25-9, 2025. Available at SSRN: <https://ssrn.com/abstract=5194301>.

²²¹ Moreover, this individual-focused explanation is critical. For a discussion of the criticism of individual-focused accountability see L. TAYLOR, *Safety in Numbers? Group Privacy and Big Data Analytics in the Developing World*, in L. TAYLOR, B. VAN DER SLOOT, L. FLORIDI (Eds.), *Group Privacy: the challenges of new data technologies*, Springer, 2017, 13-36. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2848825.

²²² It is indeed a fundamental principle of responsible AI innovation that transparency entails informing the public – and particularly those directly affected by the deployment of AI systems in law enforcement – that such systems are being, or have been, utilized. This transparency is indispensable to enable individuals to contest the outputs of these systems, to ensure accountability among those responsible, and to safeguard the full exercise of human rights. UNICRI, INTERPOL, *Responsible AI Innovation in law enforcement. AI toolkit. 02. Principles for Responsible AI Innovation*, cited, 20.

²²³ CJEU, Judgement of 7 December 2023, *OQ v Land Hessen*, Case C-634/21, ECLI:EU:C:2023:957. L. COTINO HUESO, *La primera sentencia del Tribunal de Justicia de la Unión Europea sobre decisiones automatizadas y sus implicaciones para la protección de datos y el Reglamento de inteligencia artificial*, in *Diario La Ley*, January 2024. Available at: <https://diariolaley.laleynext.es/dli/2024/01/17/la-primera-sentencia-del-tribunal-de-justicia-de-la-union-europea-sobre-decisiones-automatizadas-y-sus-implicaciones-para-la-proteccion-de-datos-y-el-reglamento-de-inteligencia-artificial>.

policing software, insofar as such scores form the basis of subsequent operations conducted by law enforcement authorities.

The question that follows is whether, given the frequent involvement of personal data in such analytical processes, the data protection framework may prove applicable and, by extension, capable of affording additional safeguards. As outlined in Chapter 1, where processing is carried out for law enforcement purposes, the applicable framework is the LED rather than the GDPR²²⁴, since the former constitutes a *lex specialis* in relation to the latter, which lays down more general rules²²⁵.

The achievement of adequate transparency within the LED framework is, however, particularly problematic²²⁶. Unlike the GDPR, the LED does not expressly enshrine transparency as a general principle²²⁷. This omission is generally explained by the specific nature of criminal justice and law enforcement activities, which necessitate differentiated levels of transparency in order to safeguard the effectiveness of investigations and protect security interests. It has been argued, nonetheless, that transparency is indirectly embedded in the principle of fairness enshrined in Article 4(1)(a)²²⁸ or, alternatively, in the data subject rights set out in Chapter III of the LED²²⁹, albeit the degree of transparency secured by the LED remains markedly weaker than that guaranteed under the GDPR²³⁰. Notwithstanding these considerations, Article 13(1) of the LED, referred to in Article 26(11) AI Act, sets out the minimum information that must be made available or provided to the data subject²³¹. This includes, *inter alia*, the purposes of the processing for which the personal data are intended; the right to lodge a complaint with a supervisory authority together with its contact details; and the existence of the right to request from the controller access to, as well as rectification or erasure of, personal

²²⁴ As outlined in Chapter 1, the Law Enforcement Directive (LED) applies when both the material and personal conditions are met, specifically, when competent authorities, such as law enforcement bodies, process personal data for law enforcement purposes. However, as also discussed in Chapter 1, the lack of conceptual clarity surrounding some of the LED's core notions complicates the distinction between the LED and the General Data Protection Regulation (GDPR), thereby blurring the boundaries between the two legal instruments.

²²⁵ The case law of the CJEU specifically clarified this statement. CJEU, Judgment of 21 June 2022, *Ligue des droits humains*, C-817/19, ECLI:EU:C:2022:491, paragraph 72. Although the level of safeguards under the two frameworks is comparable, they do not perfectly overlap. For a detailed discussion of the differences between the LED and the GDPR, see L. DRECHSLER, *Comparing LED and GDPR Adequacy: One Standard, Two Systems*, in *Global Privacy Law Review*, 1(2), 2020.

²²⁶ P. VOGIATZOGLU, T. MARQUENIE, *Assessment of the implementation of the Law Enforcement Directive*, December 2022, 33. Available at: [https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU\(2022\)740209](https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2022)740209).

²²⁷ LED, Article 4 “Member States shall provide for personal data to be: (a) processed lawfully and fairly” and GDPR, Article 5 “Personal data shall be: (a) processed lawfully, fairly and in a transparent manner in relation to the data subject (‘lawfulness, fairness and transparency’)”.

²²⁸ Commission Expert Group, Minutes of the third meeting of the Commission expert group on the Regulation (EU) 2016/679 and Directive (EU) 2016/680, 7 November 2016.

²²⁹ Article 29 Data Protection Working Party, *Opinion on some key issues of the Law Enforcement Directive (EU 2016/680)*, WP258, 29 November 2017, 3-6.

²³⁰ L. DRECHSLER, *Comparing LED and GDPR Adequacy: One Standard, Two Systems*, cited, 93–103.

²³¹ More specifically, Article 13 delineates the categories of information to be made available to the data subject: in general circumstances pursuant to paragraph (1); in particular situations pursuant to paragraph (2); and subject to specific conditions as outlined in paragraph (3).

data, or restriction of their processing. In “specific cases”²³², additional information may also be provided, such as the legal basis for the processing and, where necessary, further details, particularly in cases where personal data have been collected without the knowledge of the data subject²³³.

It has been questioned whether Article 13 LED may be assimilated to a right to notification, namely, a duty imposed on competent authorities to notify individuals affected by processing operations, once police activities are no longer jeopardised. Although this question has already been answered in the negative²³⁴, it may nonetheless be regarded as a recognised principle within the European jurisprudence, insofar as it constitutes a necessary precondition for the effective exercise of the right to an effective remedy²³⁵.

The case law of the CJEU and of the ECtHR provides important guidance in this respect²³⁶. While the ECtHR has emphasised that individuals must be informed of surveillance measures interfering with their rights under Article 8 of the ECHR once such disclosure no longer jeopardizes the investigation, the CJEU has likewise stressed that notification “*is [...] necessary to enable the persons affected to exercise, inter alia, their right to a legal remedy*”²³⁷. Indeed, effective judicial protection

²³² The provision, however, remains silent on the precise situations to which it refers. See WP29, *Opinion on some key issues of the Law Enforcement Directive (EU 2016/680)* cited, 17-18. The resolution of this ambiguity is therefore left to the Member States. Sajfert and Quintel interpret Article 13(2) as envisaging a scenario in which the competent authority proactively decides to provide information to the data subject. By contrast, Vogiatzoglou et al. regard it as establishing a complementary *ex post* right of access. See J. SAJFERT, T. QUINTEL, *Data Protection Directive (EU) 2016/680 for Police and Criminal Justice Authorities*, December 2017. Available at SSRN: <https://ssrn.com/abstract=3285873>. P. VOGIATZOGLU, K. QUEZADA TAVÁREZ, S. FANTIN, P. DEWITTE, *From Theory to Practice: Exercising the Right of Access under the Law Enforcement and PNR Directives*, in *Journal of Intellectual Property, Information Technology and E-Commerce Law*, 11(3), 2020, 274–302.

²³³ Member States may, however, pursuant to paragraph (2), delay, restrict, or altogether omit the provision of information to the data subject.

²³⁴ “*Nevertheless, the information requirements under Article 13 do not reflect such notification duty*”. P. VOGIATZOGLU, T. MARQUENIE, *Assessment of the implementation of the Law Enforcement Directive*, cited, 56.

²³⁵ F. BOEHM, P. DE HERT, *Notification, an important safeguard against the improper use of surveillance - finally recognized in case law and EU law*, in *European Journal of Law and Technology*, 2012, 3(3). F. BOEHM, P. DE HERT, *The Rights of Notification after Surveillance is over: Ready for Recognition?*, in J. BUS, M. HILDEBRANDT, M. CROMPTON, G. METAKIDES (Eds.), *Digital Enlightenment Yearbook*, 2012.

²³⁶ Such a right has been developed within the European jurisprudence. CJEU, Judgment of 21 December 2016, *Tele2 Sverige AB v Post-och telestyrelsen and Secretary of State for the Home Department v Tom Watson and Others*, C-203/15 and C-698/15, EU:C:2016:970. For a commentary on this judgment see O. POLLICINO, M. BASSINI, *La Corte di Giustizia e una trama ormai nota: la sentenza Tele2 Sverige sulla conservazione dei dati di traffico per finalità di sicurezza e ordine pubblico*, in *Diritto Penale Contemporaneo*, 2016. ECtHR, Judgment of 13 September 2018, *Big Brother Watch and Others v. the United Kingdom*, Application no. 58170/13, 62322/14 and 24960/15. “[N]otification of surveillance measures is inextricably linked to the effectiveness of remedies and hence to the existence of effective safeguards against the abuse of monitoring powers, since there is in principle little scope for any recourse by the individual concerned unless the latter is advised of the measures taken without his or her knowledge and thus able to challenge their justification retrospectively. As soon as notification can be carried out without jeopardising the purpose of the restriction after the termination of the surveillance measure, information should be provided to the persons concerned”. Similarly, Council of Europe, Committee of Ministers, *Recommendation No. R (87) 15 of the Committee of Ministers to member states regulating the use of personal data in the police sector*, 17 September 1987, Principle 2.2. For a discussion of such principle in relation to secret services see FRA, *Surveillance by intelligence services: Fundamental rights safeguards and remedies in the EU - 2023 update*, May 2023. Available at: <https://fra.europa.eu/en/publication/2023/surveillance-update>.

²³⁷ CJEU, *Tele2 Sverige AB*, cited, paragraph 121: “*Likewise, the competent national authorities to whom access to the retained data has been granted must notify the persons affected, under the applicable national procedures, as soon as that notification is no longer liable to jeopardise the investigations being undertaken by those authorities. That*

under Article 47 of the CFREU requires that individuals be made aware of interferences with their personal data, so that they may challenge the lawfulness of such measures.

A notable limitation of Article 13 LED, relevant to the present analysis, is its failure to provide explicit information on automated decision-making²³⁸. One possible avenue for addressing this lacuna lies in a combined reading of Article 11 and Recital 38, which elaborate, in the context of automated individual decision-making that “*such processing should be subject to suitable safeguards, including the provision of specific information to the data subject and the right to obtain human intervention, in particular to express his or her point of view, to obtain an explanation of the decision reached after such assessment or to challenge the decision*”²³⁹. However, this provision presents some limitations. First, Article 11 specifically refers to decisions concerning “the data subject”, thereby limiting its scope to individually targeted profiling operations and leaving collective or group profiling outside its immediate purview²⁴⁰. As discussed in Chapter 2, predictive policing may encompass both forms of profiling, depending on the specific model or methodology employed. It constitutes individual profiling when it relies on data pertaining to a particular person in order to estimate their likelihood of committing, or becoming a victim of, a crime (e.g. when systems analyse a person’s criminal record, social network, location history, or behavioural patterns to generate a risk score). In many cases, however, predictive policing is predicated on group profiling, as is most apparent in place-based models oriented towards the identification of high-risk areas rather than specific individuals and, more broadly, wherever it is employed for the in-depth analysis of data intended to inform subsequent operational decisions.

Second, questions remain regarding the extent and nature of human intervention required for a decision to fall within the scope of the provision²⁴¹, as well as whether any form of preliminary profiling would be covered²⁴². More broadly, where a decision flows from the interaction between

notification is, in fact, necessary to enable the persons affected to exercise, inter alia, their right to a legal remedy, expressly provided for in Article 15(2) of Directive 2002/58, read together with Article 22 of Directive 95/46, where their rights have been infringed (see, by analogy, judgments of 7 May 2009, Rijkeboer, C-553/07, EU:C:2009:293, paragraph 52, and of 6 October 2015, Schrems, C-362/14, EU:C:2015:650, paragraph 95)”.

²³⁸ For a discussion of the right to access in the LED see D. DIMITROVA, P. DE HERT, *The Right of Access Under the Police Directive: Small Steps Forward*, in M. MEDINA ET AL. (Eds.), *Privacy technologies and policy: 6th Annual Privacy Forum, APF 2018*. Lecture Notes in Computer Science, no. 11079, Springer International Publishing, 111-130. See also L. DRECHSLER, *Comparing LED and GDPR Adequacy: One Standard Two Systems*, cited; G. GONZÁLEZ FUSTER, *Artificial Intelligence and Law Enforcement - Impact on Fundamental Rights*, European Parliament’s Committee on Civil Liberties, Justice and Home Affairs, PE 656.295, 2020. By contrast, such information shall be given under the GDPR.

²³⁹ For a discussion on such provision see P. VOGIATZOGLOU, T. MARQUENIE, *Assessment of the implementation of the Law Enforcement Directive*, cited, 50. Notably, while the GDPR regulates automated decision-making as part of the rights afforded to data subjects, the LED incorporates it within its general principles, reflecting a structural divergence between the two instruments.

²⁴⁰ VOGIATZOGLOU, MARQUENIE, *Assessment of the implementation of the Law Enforcement Directive*, cited, 50

²⁴¹ O. LYNSEY, *Criminal Justice Profiling and EU Data Protection Law: Precarious Protection from Predictive Policing*, in *International Journal of Law in Context*, 15(2), 2019, 162–176.

²⁴² SAJFERT, QUINTEL, *Data Protection Directive (EU) 2016/680 For Police and Criminal Justice Authorities*, cited.

algorithmic output and human intervention, the applicability of Article 11 to such a process remains an open question. Equally significant, in this regard, is the degree of consistency – or lack thereof – with which Member States have transposed this provision²⁴³. The limitations examined above reveal that the LED falls short of adequately equipping the individual concerned with the means necessary to effectively access or make use of the information to which they are entitled as a data subject²⁴⁴, with the consequence that no effective additional safeguards appear to be afforded.

Beyond access to judicial remedies and the possibility of lodging a complaint with the market surveillance authority²⁴⁵, the AI Act does not confer upon affected individuals a right to obtain human review of an adverse decision adopted by a deployer²⁴⁶. Moreover, the pursuit of compensation is further complicated by the absence of explicit provisions on liability within the legislative framework²⁴⁷. Nevertheless, the AI Act operates without prejudice to the competences, tasks, powers, and independence of national authorities or bodies charged with supervising the application of Union law safeguarding fundamental rights²⁴⁸.

²⁴³ VOGIATZOGLOU, MARQUENIE, *Assessment of the implementation of the Law Enforcement Directive*, cited.

²⁴⁴ In general, the right of access is regarded as a prerequisite for the effective exercise of all other data protection rights (right to rectification, erasure, and restriction of processing). This view is supported in the literature by Dimitrova and De Hert, who describe it as a foundational right enabling individuals to understand and challenge the processing of their personal data. Similarly, Vogiatzoglou et al. emphasize its practical significance under both the Law Enforcement and PNR Directives, while Quezada Tavárez highlights its role in recalibrating the balance between security and fundamental rights through the notion of “informational power”. See DIMITROVA, DE HERT, *The Right of Access Under the Police Directive: Small Steps Forward*, cited; VOGIATZOGLOU ET AL., *From Theory to Practice: Exercising the Right of Access under the Law Enforcement and PNR Directives*, cited; K. QUEZADA TAVÁREZ, *Impact of the Right of Access on the Balance between Security and Fundamental Rights: Informational Power as a Tool to Watch the Watchers*, in *European Data Protection Law Review*, 7(1), 2021, 59–73. See, inter alia, CJEU, Judgment of 7 May 2009, *College van burgemeester en wethouders van Rotterdam v M.E.E. Rijkeboer*, Case C-553/07, EU:C:2009:293, paragraphs 51–52; Judgment of 20 December 2017, *Peter Nowak v Data Protection Commissioner*, Case C-434/16, EU:C:2017:994, paragraph 57. The right of access is also recognised as an element of the fundamental right to the protection of personal data under Article 8(2) of the CFREU.

²⁴⁵ AI Act, Article 85. Some have claimed that this absence conflicts with the CJEU’s case law. T. MARQUENIE, *The Police and Criminal Justice Authorities Directive: Data Protection Standards and Impact on the Legal Framework*, in *Computer Law & Security Review*, 33(3), 2017, 324–340; M. LEISER, B. CUSTERS, *The Law Enforcement Directive: Conceptual Challenges of EU Directive 2016/680*, in *European Data Protection Law Review*, 5(3), 2019, 367–378.

²⁴⁶ An example in this regard is Recital 71 of the GDPR. This alternative is more accessible to affected individuals, as it is both less time-consuming and less costly.

²⁴⁷ In October 2020, acting under Article 225 TFEU, the European Parliament adopted a legislative own-initiative resolution on civil liability for artificial intelligence, inviting the European Commission to submit a corresponding legislative proposal. In response, on 28 September 2022, the Commission presented the Proposal for an Artificial Intelligence Liability Directive (AILD), thereby delivering on the objectives set out in the *White Paper on Artificial Intelligence* and addressing the Parliament’s request. The legislative process, however, was later discontinued, as the European Commission withdrew the proposal from further examination.

²⁴⁸ AI Act, Article 77.

Part II: Algorithmic fairness under the AI Act: implications for law enforcement

1. Algorithmic fairness

One of the central themes in the literature on AI risks is the one concerning bias, that is, situations where an AI system systematically and unfairly favours certain individuals or groups over others or demonstrates significantly better performance for some groups than for others²⁴⁹. Race and gender are among the most extensively discussed dimensions along which biases in AI systems can be identified. In Chapter 2, I discussed how face recognition technologies, for example, tend to perform worse for underrepresented groups, particularly for dark-skinned women. However, it is evident that AI systems – particularly those deployed within the public sector – must operate in a fair manner²⁵⁰. The stakes are particularly high for AI systems employed in law enforcement, where algorithmic unfairness poses profound threats to fundamental rights, potentially culminating in wrongful prosecutions and unjustified deprivations of liberty²⁵¹.

The literature identifies several forms of bias²⁵², which are: implicit bias, referring to the tendency to make assumptions based on personal experiences and mental models that may not hold more generally; selection bias, or the tendency to choose datasets or features that are not representative of real-world distributions; group attribution bias, denoting the inclination to generalize characteristics or behaviours observed in some members of a group to the entire group; reporting bias, which arises when the frequency of events, properties, or outcomes recorded in a dataset does not accurately reflect their real-world occurrence. These forms of bias give rise to potential sources of “unfairness” in algorithmic systems²⁵³, which may be identified and categorised as follows²⁵⁴:

²⁴⁹ Scholars have long debated this topic, as evidenced by the work of B. FRIEDMAN, H. NISSENBAUM, *Bias in computer systems*, in *ACM Transactions on Information Systems*, 14(3), 1996, 330-347. FRA, *Bias in algorithms - Artificial intelligence and discrimination*, December 2022. Available at: <https://fra.europa.eu/en/publication/2022/bias-algorithm>.

²⁵⁰ A. JOBIN, M. IENCA, E. VAYENA, *The global landscape of AI ethics guidelines*, in *Nature Machine Intelligence*, 1, 2019, 389-399. Available at: <https://doi.org/10.1038/s42256-019-0088-2>. HLEG, *Ethics guidelines for trustworthy AI*, cited, 12.

²⁵¹ UNICRI, INTERPOL, *Responsible AI Innovation in law enforcement. AI toolkit. 02. Principles for Responsible AI Innovation*, cited, 23.

²⁵² UNICRI, INTERPOL, *Responsible AI Innovation in law enforcement. AI toolkit - 07. Technical Reference Book*, cited, 44-45.

²⁵³ S. MOUSSAWI, X. DENG, K. D. JOSHI, *AI and Discrimination: Sources of Algorithmic Biases*, in *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 55(4), 2024, 6-11. Available at: <https://doi.org/10.1145/3701613.370161>.

²⁵⁴ J. KLEINBERG ET AL., *Discrimination in the Age of Algorithms*, 25548 NBER Working Paper Series, 2019, 21-23; S. BAROCAS, A.D. SELBST, *Big Data's Disparate Impact*, in *104 California law review*, 671, 2016, 677-680. P. HACKER, *Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law*, in *Common market law review*, 2018, 1146-1148.

- (1) Problem definition²⁵⁵: unfairness may arise at the initial stage of designing the model, when the objectives, success metrics, or problem formulations reflect subjective or biased assumptions;
- (2) Training data²⁵⁶: bias can be introduced through the datasets used to train the model, which may contain historical inequalities, underrepresentation of certain groups or measurement errors²⁵⁷;
- (3) Selection of target features²⁵⁸: the choice of which attributes are considered relevant to the algorithm's decision-making process is extremely delicate and can perpetuate structural biases if these features correlate with protected characteristics;
- (4) Use of proxy data²⁵⁹: even when sensitive attributes such as gender, race, or age are excluded, other variables may act as "proxies" and indirectly reproduce discriminatory effects²⁶⁰;
- (5) Intentional discrimination²⁶¹: unfairness may result from deliberate choices by developers, who can embed discriminatory rules or design features within the model's architecture to achieve particular biased outcomes.

All of the foregoing sources of unfairness inevitably impinge upon the outcomes generated by an AI system. As the literature affirms, however, such outcomes constitute only one component of the broader decision-making process²⁶², which is structured around multiple distinct stages²⁶³: scoring, classification, and decision-making. The first stage, scoring, consists of assigning a score to an entity, that is, determining the likelihood that the entity possesses the predicted property. For example, if a system is employed to predict the risk of crime, the score may represent the probability that an individual will commit an offence. As scholars emphasise, this first stage should be primarily grounded in epistemic considerations, specifically, in obtaining scores that most accurately reflect the likelihood of the target property being present. The second stage, classification, functions as a bridge between scoring and decision-making, providing triggers for action. By classifying entities within

²⁵⁵ KLEINBERG ET AL., *Discrimination in the Age of Algorithms*, cited, 21-23; BAROCAS, SELBST, *Big Data's Disparate Impact*, cited, 677-680.

²⁵⁶ *Ibidem*. HACKER, *Teaching fairness to artificial intelligence*, cited, 1146-1148.

²⁵⁷ BAROCAS, SELBST, *Big Data's Disparate Impact*, cited, 683. On the importance of the quality of the data used see B. HANSON, S. STALL, J. CUTCHER GERSHENFELD, K. VROUWENVELDER, C. WIRZ, Y. RAO & G. PENG, *Garbage in, garbage out: mitigating risks and maximizing benefits of AI in research*, in *Nature*, 623(2), 2023, 28-31. See also D. HELLMAN, *Algorithmic Fairness*, in E.N. ZALTA, U. NODELMAN (Eds.), *The Stanford Encyclopedia of Philosophy*, 2025. Available at: <https://plato.stanford.edu/archives/fall2025/entries/algorithmic-fairness/>.

²⁵⁸ BAROCAS, SELBST, *Big Data's Disparate Impact*, cited, 677-680.

²⁵⁹ *Ibidem*. HELLMAN, *Algorithmic Fairness*, cited.

²⁶⁰ Due to the fact that proxy data often correlate closely with protected characteristics. For example, zip codes are generally considered a proxy for race. J. GERARDS, R. XENIDIS, *Algorithmic discrimination in Europe: challenges and opportunities for gender equality and non-discrimination law*, Publications Office, 2021, 44.

²⁶¹ *Ibidem*.

²⁶² F. LAGIOIA, R. ROVATTI, G. SARTOR, *Algorithmic fairness through group parities? The case of COMPAS-SAPMOC*, in *AI & Society*, 38, 2023, 459-478. Available at: <https://doi.org/10.1007/s00146-022-01441-y>. Indeed, the process of decision-making encompasses the articulation of objectives, the identification of pertinent ethical and legal constraints, the systematic assessment of the potential consequences of various courses of action, and the selection of the alternative most consistent with the predefined objectives and constraints.

²⁶³ F. LAGIOIA ET AL., *Algorithmic fairness through group parities?*, cited, 460-462.

certain score intervals (for instance, as having a high, medium, or low likelihood of possessing the target feature), the classification process anticipates how such entities are likely to be treated. For example, in relation to AI systems used for risk assessment, classifying all individuals who obtain a score of 70% as high-risk may have significant implications for the way they are subsequently treated. As is evident, neither threshold setting nor the subsequent classification processes are neutral, and both require value judgments. Such decisions inevitably carry significant consequences. As discussed in Chapter 2, all entities that fall above the established threshold are regarded as possessing the targeted feature, even though only some of them in fact do (*i.e.* “true positives”). Conversely, entities falling below the threshold are deemed not to possess the targeted feature, although only some of them genuinely do not (*i.e.* “true negatives”). The final stage is decision-making, typically entrusted to a human decision-maker, which involves determining and implementing the appropriate course of action once the classification has been completed. For instance, a police officer may take into account a risk assessment and decide to place an individual under stricter surveillance, which is a detrimental treatment.

The concept commonly invoked in the literature concerning fairness in algorithmic systems is “algorithmic fairness”²⁶⁴, which has been broadly defined as encompassing “*debates in the literature regarding when and whether the use of machine learning algorithms in different ways and in different contexts is fair*”²⁶⁵. Despite the extensive scholarly engagement with the subject, addressing the issue remains particularly complex, as the notion of fairness is examined from markedly diverse theoretical and disciplinary perspectives²⁶⁶. Indeed, scholars tend to emphasise different dimensions of fairness – as distributive or procedural justice²⁶⁷, as statistical or social fairness, as individual or group fairness²⁶⁸ – thereby hindering the formulation of a coherent and universally accepted definition or operational framework. This conceptual plurality, in turn, exacerbates the difficulty of translating abstract principles of fairness into concrete, enforceable rules governing the design and deployment of AI systems.

²⁶⁴ S. MITCHELL, E. POTASH, S. BAROCAS, A. D’AMOUR, K. LUM, *Algorithmic Fairness: Choices, Assumptions, and Definitions*, in *Annual Review of Statistics and Its Application*, 8, 2021, 141-163.

²⁶⁵ For a detailed discussion on algorithmic fairness see HELLMAN, *Algorithmic Fairness*, cited. In particular, the literature dealing with algorithmic fairness widely increased after a particular case that involved the American Justice, namely the use of the algorithmic tool COMPAS to establish the risk of reoffending. For a discussion of this algorithmic tool see LAGIOIA ET AL., *Algorithmic fairness through group parities?*, cited, 462.

²⁶⁶ HELLMAN, *Algorithmic Fairness*, cited. J. GERARDS, R. XENIDIS, *Algorithmic discrimination in Europe: challenges and opportunities for gender equality and non-discrimination law*, cited; R. XENIDIS, L. SENDEN, *EU Non-discrimination Law in the Era of Artificial Intelligence: Mapping the Challenges of Algorithmic Discrimination*, in U. BERNITZ ET AL. (Eds.), *General Principles of EU Law and the EU Digital Order*, Kluwer Law International, 2020, 151-182.

²⁶⁷ VARSHNEY, *Trustworthy Machine Learning*, cited, 131.

²⁶⁸ *Ibidem*, 132.

To illustrate this point, it is instructive to consider the studies that have examined the COMPAS algorithm used in the U.S. criminal justice system. As discussed in Chapter 2²⁶⁹, the investigation conducted by ProPublica revealed that the software was affected by significant racial bias²⁷⁰. In the same year, however, researchers affiliated with the company responsible for commercialising COMPAS published a rejoinder to the ProPublica report, seeking to substantiate the system's "accuracy equity" and "predictive parity"²⁷¹. A noteworthy contribution addressed this issue by analysing two simplified versions of the COMPAS algorithm – designated SAPMOC I and SAPMOC II – with the aim of evaluating the fairness of the algorithmic tool against various fairness criteria, referred to by the authors as "group-parity metrics"²⁷². The chosen criteria²⁷³ include statistical parity²⁷⁴, conditional procedure accuracy equality/equality of opportunity²⁷⁵, calibration/conditional use accuracy equality²⁷⁶, false rate/conditional use error²⁷⁷ and treatment equality²⁷⁸. The study found that in relation to the two SAPMOCs, while some fairness criteria are satisfied²⁷⁹, others are violated²⁸⁰. Therefore, although they may appear unbiased at first glance, they nonetheless breach several key fairness measures²⁸¹. These violations largely stem from differences in base rates between the two groups: specifically, one group includes a higher proportion of individuals with a criminal history, which makes it impossible to satisfy all fairness criteria simultaneously. Such disparities in the base rates may arise, for example, from biases in the input data: if a particular group is subjected

²⁶⁹ J. ANGWIN, J. LARSON, S. MATTU, L. KIRCHNER, *Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks*, ProPublica, 23 May 2016. Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

²⁷⁰ Specifically, "[o]n one hand, black non-recidivists were more likely than white non-recidivists to be erroneously subjected to the detrimental consequences linked to a recidivism prediction. On the other hand, white recidivists were more likely than black recidivists to erroneously obtain the more favorable treatment for expected non-recidivists". *Ibidem*.

²⁷¹ W. DIETERICH, C. MENDOZA, T. BRENNAN, *COMPAS risk scales: demonstrating accuracy equity and predictive parity*, Northpoint Inc 7 (7.4), 1, 2016.

²⁷² For an interesting discussion on COMPAS and fairness see D. HELLMAN, *Algorithmic Fairness*, cited.

²⁷³ The literature has identified multiple statistical measures of algorithmic fairness. Among others, B. HEDDEN, *On Statistical Criteria of Algorithmic Fairness*, in *Philosophy & Public Affairs*, 49(2), 2021, 214–218.

²⁷⁴ Statistical parity requires the proportion of positive (recidivism) and negative (no recidivism) predictions to be equal in each group. LAGIOIA ET AL., *Algorithmic fairness through group parities?*, cited, 466.

²⁷⁵ According to conditional procedure accuracy equality, the members in each group who exhibit the same behaviour should be treated equally in equal proportion. *Ibidem*.

²⁷⁶ According to calibration, the proportion of correct predictions should be equal for each class within each group. *Ibidem*, 467.

²⁷⁷ It measures the proportion of erroneous predictions for each class relative to the total predictions.

²⁷⁸ According to treatment equality, the ratio between errors in positive and negative predictions should be equal across all groups.

²⁷⁹ In relation to SAPMOC I calibration/conditional use accuracy equality and false rate/conditional use error equality.

²⁸⁰ In relation to SAPMOC I statistical parity, conditional procedure accuracy equality, and treatment equality.

²⁸¹ This has been highlighted also by J. KLEINBERG, S. MULLAINATHAN, M. RAGHAVAN, *Inherent trade-offs in the fair determination of risk scores*, 2016, 3. Available at: arXiv:1609.05807v2. A. CHOULDECHOVA, *Fair prediction with disparate impact: a study of bias in recidivism prediction instruments*, in *Big Data*, 5(2), 2017, 153–163.

to more intensive monitoring or control – often as a result of historical prejudices and structural inequalities – instances of criminal behaviour are detected more frequently within that group²⁸².

The contribution also provides valuable insights in relation to threshold setting. Raising the threshold – while leaving the underlying scoring mechanism unchanged – reduces false positives but increases false negatives, and vice versa. Similarly, the application of different thresholds across groups may enable the fulfilment of specific fairness criteria (e.g., statistical equality), yet it can also undermine other fairness dimensions or diminish the predictive accuracy of the system²⁸³. In general, the appropriateness of adjusting the threshold depends on the relative costs or harms associated with different types of misclassifications. For example, in the context of COMPAS, the cost of erroneously classifying a person as a recidivist (and thus, for instance, denying parole) must be weighed against the cost of erroneously classifying that person as non-recidivist. Such judgments inevitably rest on value-based considerations, which should be made by the decision-maker, not by the private company developing or marketing the software.

Inevitably, divergent assessments of fairness may arise depending on which normative criteria are selected and applied, a tension that has been extensively theorised and documented in the rapidly expanding literature on fairness in machine learning²⁸⁴. The coexistence of multiple and often conflicting fairness criteria implies that, except in rare or exceptional cases, it is impossible to treat two groups of individuals equally according to all fairness measures simultaneously²⁸⁵. The main reason is that the concept of fairness as defined in statistical terms does not necessarily correspond to the concept of fairness in legal or ethical terms. Thus, any meaningful evaluation of an algorithmic

²⁸² M. OSWALD, A. BABUTA, *Data analytics and algorithmic bias in policing*, Royal United Services Institute for Defence and Security Studies, 2019. Available at: https://assets.publishing.service.gov.uk/media/5d7f6b2540f0b61ccd4a4b80/RUSI_Report_-_Algorithms_and_Bias_in_Policing.pdf.

²⁸³ Setting different thresholds for different groups entails treating individuals with identical scores differently solely on the basis of group membership. Such differential treatment may, in certain circumstances, be justified, for instance, to correct biases in input data or feature selection, or to promote broader policy objectives such as affirmative action or social equity. However, this approach may at the same time result in the violation of other fairness criteria. Moreover, the appropriateness of applying group-specific thresholds is highly context-dependent. For example, in domains such as law enforcement, the use of group-specific thresholds is generally regarded as unjustified, whereas in contexts like algorithmic disease prediction it may be considered acceptable or even necessary to ensure equitable outcomes. Adjusting thresholds uniformly across all groups can mitigate certain harms, such as excessive detention, but may simultaneously compromise overall accuracy. LAGIOIA ET AL., *Algorithmic fairness through group parities?*, cited, 473-474.

²⁸⁴ R. BERK, H. HEIDARI, S. JABBARI, M. KEARNS, A. ROTH, *Fairness in criminal justice risk assessments: the state of the art*, in *Sociological Methods & Research*, 50(1), 2018, 3-44.

²⁸⁵ J. KLEINBERG ET AL., *Inherent trade-offs in the fair determination of risk scores*, cited. CHOULDECHOVA, *Fair prediction with disparate impact*, cited. The conclusions reached by the authors are referred to as the “impossibility result” or the “Kleinberg-Chouldechova impossibility theorem”. These conclusions have, however, been challenged by Green and Hu, who dispute the broader significance of the so-called “impossibility result” within the discourse on algorithmic fairness. As they observe, “*labeling a particular incompatibility of statistics as an impossibility of fairness generally is mistaking the map for the territory*”. B. GREEN, L. HU, *The Myth in the Methodology: Towards a Recontextualization of Fairness in Machine Learning*, Presented at the Machine Learning: The Debates workshop at the 35th International Conference on Machine Learning, Stockholm, Sweden. Available at: <https://www.benzevgreen.com/wp-content/uploads/2019/02/18-icmldebates.pdf>.

tool's fairness should begin with a prior clarification of the very concept of fairness itself²⁸⁶. For some scholars, for instance, fairness means “*treating like cases alike*” thereby conceiving fairness as a comparative notion. This understanding focuses on how one person or group is treated in relation to another, emphasising equality of treatment across comparable situations. Other scholars, however, view fairness as a non-comparative concept, concerned not with relative treatment but with whether an individual or group is treated as they ought to be treated. The basis of such a complaint may take several forms. It might point to the fact that an algorithm is inaccurate, or less accurate than it could or should be; alternatively, it may object that an individual judgment is impermissibly based on group characteristics rather than individual merit. Finally, some argue that the use of algorithms itself in making consequential decisions that affect people is inherently unfair, as it replaces human judgment with machine-based decision-making. The notion of fairness, however, is inherently context-dependent and may differ depending on the objectives and intended function of the AI system²⁸⁷.

2. The legal principle of non-discrimination

From a legal standpoint, decisions based on AI systems' outputs have been evaluated through the legal principle of non-discrimination²⁸⁸ and the lens of anti-discrimination law²⁸⁹. Within the EU anti-discrimination framework, a distinction is drawn between direct discrimination²⁹⁰, which involves differential treatment based on a protected ground – such as sex, racial or ethnic origin, religion, disability, age, or sexual orientation – and indirect discrimination, which arises when an apparently neutral provision, criterion, or practice disproportionately disadvantages individuals on such

²⁸⁶ To cite some literature, see E. BENJAMIN, *Algorithmic Fairness and Base Rate Tracking*, in *Philosophy & Public Affairs*, 50(2), 2022, 239–266; B. BABIC, Z. JOHNSON KING, *Algorithmic Fairness and Resentment*, in *Philosophical Studies*, 182(1), 2025, 87–119; S. LAZAR, J. STONE, *On the Site of Predictive Justice*, *Noûs*, 58(3), 2024, 730–754.

²⁸⁷ Deborah Hellman makes the example of COMPAS, stating that the fact that the algorithm generates more false positives for black inmates than for white inmates seems unfair to the black inmates if the result will be incarceration but not if it will be beneficial services. HELLMAN, *Algorithmic Fairness*, cited.

²⁸⁸ The principle of non-discrimination is enshrined in Articles 2 and 3(3) TEU. In addition, it is embedded across EU policies through Articles 10 and 19 TFEU and Article 21 of the CFREU. Right to equality and non-discrimination is specifically a human right that law enforcement authorities are legally obliged to respect. United Nations Human Rights Committee, General comment No. 18: Non-discrimination, Thirty-seventh session, 1989. Such rights impose on States and State bodies the obligation to ensure that individuals are treated equally and are equally able to exercise and enjoy all their rights, and that everyone is protected against direct or indirect discrimination based on “protected characteristics” such as gender, race, ethnic origin, age, religion, ability, and sexual orientation.

²⁸⁹ Directives 2006/54/EC for the ground of sex (the Gender Recast Directive); Directive 2000/43/EC for the ground of race or ethnic origin (the Race Equality Directive) and Directive 2000/78/EC for the grounds of disabilities, religion or belief, age and sexual orientation (the Employment Equality Directive). J. GERARDS, R. XENIDIS, *Algorithmic discrimination in Europe: challenges and opportunities for gender equality and non-discrimination law*, cited.

²⁹⁰ The European legal framework considers direct discrimination when “one person is treated less favourably than another is, has been or would be treated in a comparable situation’ on the basis of one of the protected grounds defined in the relevant directives”. Article 2(2)(a) of Directive 2000/43/EC; Article 2(2)(a) of Directive 2000/78/EC; Article 2(a) of Directive 2004/113/EC and Article 2(1)(a) of Directive 2006/54/EC.

grounds²⁹¹. While in the former case the detrimental decision is directly based on protected grounds, in the latter case the discriminatory outcome is indirect, placing persons of a particular racial or ethnic origin at a disadvantage compared with others, unless that provision, criterion, or practice is objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary²⁹².

Anti-discrimination law remains fully applicable in contexts where AI systems are deployed²⁹³. Direct discrimination may extend to forms of algorithmic discrimination, given that it does not require proof of intent. Direct discrimination may also encompass cases where protected characteristics are explicitly included as relevant variables within an AI model and treated as negative factors, rendering such practices unlawful²⁹⁴. Furthermore, the concept of discrimination by association, developed by the CJEU²⁹⁵, strengthens protection by applying when an individual does not personally possess a protected characteristic but suffers unfavourable treatment due to his/her association with a protected group. Similarly, direct discrimination provisions may apply in cases of algorithmic misclassification, where an AI system erroneously assumes that an individual possesses a protected characteristic²⁹⁶. However, the integration of AI systems into decision-making processes significantly complicates the application of traditional legal principles and rules²⁹⁷. For instance, the use of proxy data – variables that indirectly reflect protected characteristics – can obscure the presence of discrimination, allowing

²⁹¹ R. XENIDIS, L. SENDEN, *EU Non-discrimination Law in the Era of Artificial Intelligence*, 19. The authors examine the challenges that advances in artificial intelligence raise in relation to the EU principle of non-discrimination. Conversely, in the context of U.S. constitutional law, issues regarding whether an individual receives treatment equal to that of others fall within the scope of the *Equal Protection Clause*, while questions concerning whether an individual is treated as they ought to be, according to a non-comparative standard, are governed by the *Due Process Clause*.

²⁹² For a discussion on the differences between direct and indirect discrimination see J. MALISZEWSKA-NIENARTOWICZ, *Direct and Indirect Discrimination in European Union Law – How to Draw a Dividing Line?*, in *International Journal of Social Sciences*, III(1), 2014, 41-55. As far as indirect discrimination is concerned see Article 2(2)(b) Directive 2000/43/EC; Article 2(2)(b) Directive 2000/78/EC; Article 2(b) Directive 2004/113/EC; Article 2(1)(b) Directive 2006/54/EC.

²⁹³ The AI Act does not affect anti-discrimination legislation. See AI Act, Recital 45 which states that “*Practices that are prohibited by Union law, including data protection law, non-discrimination law, consumer protection law, and competition law, should not be affected by this Regulation*”.

²⁹⁴ XENIDIS, SENDEN, *EU Non-discrimination Law in the Era of Artificial Intelligence*, cited, 19.

²⁹⁵ CJEU, Judgment of 17 July 2008, *S. Coleman v Attridge Law and Steve Law*, Case C-303/06, EU:C:2008:415; CJEU, Judgment of 16 July 2015, ‘*CHEZ Razpredelenie Bulgaria*’ AD v *Komisija za zashtita ot diskriminatsia*, Case C-83/14 EU:C:2015:480.

²⁹⁶ OLIVATO, *The AI Act: regulatory and technical challenges*, cited, 282.

²⁹⁷ See F. ZUIDERVEEN BORGESIU, *Discrimination, artificial intelligence and algorithmic decision-making*, commissioned by the Council of Europe’s European Commission Against Racism and Intolerance (ECRI), 2018. Available at: <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>. I. BARTOLETTI, R. XENIDIS, *Study on the impact of artificial intelligence systems, their potential for promoting equality – including gender equality - and the risks they may cause in relation to non-discrimination*, commissioned by the Committee on Anti-Discrimination, Diversity and Inclusion (CDADI) and the Gender Equality Commission (GEC) of the Council of Europe, 2023. Available at: <https://edoc.coe.int/en/artificial-intelligence/11649-study-on-the-impact-of-artificial-intelligence-systems-their-potential-for-promoting-equality-including-gender-equality-and-the-risks-they-may-cause-in-relation-to-non-discrimination.html>. C. NARDOCCI, *Artificial Intelligence-based Discrimination: Theoretical and Normative Responses. Perspectives from Europe*, in *DPCE Online*, 60(3), 2023, 2367–2394. Available at: <https://doi.org/10.57660/dpceonline.2023.1981>.

biased outcomes to go undetected. These situations are better examined through the lens of indirect discrimination²⁹⁸, as this framework addresses the disparate effects of practices that appear neutral on their face but produce disproportionate disadvantages for certain individuals or groups of individuals²⁹⁹.

Additional challenges arise from the reliance of AI models on multiple statistically correlated variables, which makes it difficult to isolate any single protected attribute as the decisive factor influencing the outcome. Moreover, intersectionality compounds these difficulties. Indeed, algorithmic outputs frequently reflect unique combinations of attributes (such as gender, ethnicity, disability status, or sexual orientation), creating patterns of disadvantage that do not neatly fit within the single-ground approach traditionally adopted by anti-discrimination law³⁰⁰. Finally, the opacity and complexity of AI systems further intensify these concerns. The lack of transparency in algorithmic processes and the often-intricate decision-making chains make it exceedingly difficult to determine whether an AI system's output is discriminatory, thereby posing significant challenges to effective legal scrutiny and enforcement. This Thesis will not explore how anti-discrimination law protects individuals from algorithmic discrimination, as such analysis falls beyond its primary scope. Nevertheless, the existing literature on this topic is already extensive³⁰¹.

3. The AI Act's provisions on bias and non-discrimination

The AI Act put much importance to the principle of non-discrimination³⁰² and seeks to address concerns related to bias by recognising that discrimination constitutes a potential risk associated with AI systems, specifically high-risk AI systems, one that must be identified, assessed and mitigated. In addition, the Regulation places strong emphasis on data quality and governance, acknowledging that

²⁹⁸ On the limitations of indirect administration see GERARDS, XENIDIS, *Algorithmic discrimination in Europe: challenges and opportunities for gender equality and non-discrimination law*, cited, 73

²⁹⁹ R. XENIDIS, L. SENDEN, *EU Non-discrimination Law in the Era of Artificial Intelligence*, cited. GERARDS, XENIDIS, *Algorithmic discrimination in Europe: challenges and opportunities for gender equality and non-discrimination law*, cited, 72.

³⁰⁰ Although the Court of Justice has acknowledged the existence of multiple discrimination, it has consistently refrained from recognising intersectional discrimination as an autonomous category within the framework of EU equality law. CJEU, Judgment of 24 November 2016, *David L. Parris v. Trinity College Dublin and Others*, Case C-443/15, EU:C:2016:897.

³⁰¹ ZUIDERVEEN BORGESIU, *Discrimination, artificial intelligence and algorithmic decision-making*, cited; FRA, *Bias in algorithms*, cited.

³⁰² The AI Act makes repeated reference to the principle of non-discrimination. It explicitly draws upon the *Ethics Guidelines for Trustworthy AI*, which articulate diversity, non-discrimination, and fairness as key ethical principles guiding the development and deployment of AI systems. In the context of prohibited practices, the AI Act underscores that such practices are contrary to Union values and to the fundamental rights enshrined in the CFREU, most notably, the right to non-discrimination. The principle is further reaffirmed in several recitals, including Recitals 48, 58, 60, 68, and 142.

the quality of the data is essential to prevent discriminatory outcomes³⁰³. Human oversight throughout the lifecycle of AI systems also represents a requirement that is aimed at minimising the risk of algorithmic discrimination and, more generally, at reducing the detrimental effects of AI systems.

3.1.Discrimination as a potential risk of using AI systems: the central role of a Fundamental Rights Impact Assessment

In 2022, the ALIGNER project conducted a survey among practitioners and professionals operating within the field of law enforcement and policing, with the aim of assessing whether, and to what extent, AI systems could enhance operational capabilities in this domain³⁰⁴. The findings revealed that 76% of law enforcement authorities refrain from employing AI in this context, owing to a combination of societal, legal, and ethical challenges, with nearly half of respondents (49%) identifying fundamental rights concerns and ethical constraints as the primary impediment.

These findings underscore the imperative that, prior to the deployment of AI systems in the exercise of their functions, law enforcement authorities must undertake a comprehensive process of risk identification, assessment, and mitigation. Specifically, law enforcement authorities deploying AI systems must undertake a systematic risk management process comprising two distinct phases. First, they must conduct comprehensive risk assessment, involving the identification, analysis, and evaluation of potential risk that may emerge from AI deployment. This assessment phase requires agencies to map possible adverse outcomes, analyse their likelihood and severity, and evaluate their significance in relation to established risk tolerance thresholds. Second, having identified and evaluated these risks, authorities must implement appropriate technical and organisational measures, selecting and deploying mitigation strategies proportionate to the identified threats.

From a legal perspective, Article 27 of the AI Act requires law enforcement authorities to conduct a Fundamental Rights Impact Assessment (hereinafter “FRIA”)³⁰⁵. Impact assessments are, in general,

³⁰³ One of the most interesting contributions concerning the relationship between data and bias is provided by B. HANSON, ET AL., *Garbage in, garbage out*, cited. A subsequent work is S. MAYSON, *Bias In, Bias Out*, University of Georgia School of Law Legal Studies Research Paper, No 2018-35.

³⁰⁴ *Artificial Intelligence Roadmap for Policing and Law Enforcement* project, financed under the Horizon 2020 programme. Further information about the project is available at: <https://cordis.europa.eu/project/id/101020574>.

³⁰⁵ The FRIA is not an isolated or novel instrument; rather, it forms part of a broader and well-established tradition of *ex ante* evaluation tools designed to anticipate and mitigate negative consequences of policies, technologies, or practices. The FRIA follows the path of earlier models such as Human Rights Impact Assessments (HRIA), Social Impact Assessments (SIA), and Ethical Impact Assessments (EIA). In addition, there are impact assessments tailored to the protection of specific rights, most notably the Data Protection Impact Assessment (DPIA). For an interesting discussion on FRIA in the AI Act see A. MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template*, in *Computer Law & Security Review*, 2024, 1-18. Available at: <https://www.sciencedirect.com/science/article/pii/S0267364924000864>.

tools for “*weighting fundamental rights*”³⁰⁶. The contemporary understanding of society as a “risk society” has prompted legislators to introduce an expanding range of *ex ante* impact assessment mechanisms. These instruments require public and private decision-makers to evaluate, prior to implementation, the potential effects of their measures – both in terms of means and outcomes – on the fundamental rights and interests at stake³⁰⁷. As discussed in Chapter 1, the tension between the protection of fundamental rights and the overarching imperatives of crime prevention, investigation, detection, prosecution, and the maintenance of public order and security is inherent and unavoidable³⁰⁸. In fact, the pursuit of these objectives necessarily entails the imposition of certain limitations upon individual rights. At both the European and national levels, rights enshrined in the Constitutions or in the Charters of Rights are not absolute and may, under specific conditions, be lawfully restricted. Therefore, rights can be assigned varying degrees of importance and their possible compromise accepted or not. Nevertheless, any such interference must be subject to careful assessment and appropriate safeguards to ensure that the balance between individual rights and collective interests is duly maintained.

To this end, the FRIA, whose primary objective is to detect potential interferences with the rights of individuals or groups and to outline appropriate safeguards and remedial actions in the case that such

³⁰⁶ E.C. DONET, *The Fundamental Rights Impact Assessment By Deployers Of Artificial Intelligence Systems In The Regulation*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 524. It is well known that it is the Constitution, as well as the CFREU that, in the first instance, establishes a hierarchy among rights by conferring upon certain rights the status of fundamental rights. Nevertheless, constitutional frameworks may also recognise a number of rights as being placed on an equal normative footing. In such cases, it falls to the legislator to determine, within specific statutory provisions, which rights are to be accorded prevalence or priority in particular contexts. In the majority of situations, however, the determination of which right should prevail is not expressly made by the legislator. Rather, this responsibility lies with the public administration, which is required to perform a balancing exercise, guided by the principles of proportionality and reasonableness, to reconcile competing rights and interests within the framework of constitutional and administrative law. In the final instance, the judiciary is entrusted with reviewing the conformity of such decisions with the constitutional framework and the rule of law. Various methodologies have been developed to conduct impact assessments, each based on different criteria. These may be distinguished according to the potential consequences, for instance, the rights and legitimate interests that may be affected; the type of technology employed; the degree of third-party involvement, such as that of providers, customers, sellers, or auditors; or the primary risks identified, including risks of misinformation, discrimination, or defencelessness. Mantelero advocates for a hybrid model that integrates human rights, ethical, and social impact assessments. Several countries have already developed specific methodologies and tools for conducting algorithmic impact assessments. Notable examples include the Netherlands’ *Fundamental Rights and Algorithm Impact Assessment*, the *Canadian Algorithmic Impact Assessment Tool*, and the *U.S. Government’s Algorithmic Impact Assessment Tool*.

³⁰⁷ A. MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act*, cited, 13 states that “as a consequence of the transformation of modern society into a risk society, or at least a society in which many activities involve exposure to risks and which is characterised by the emergence of new risks”.

³⁰⁸ As several authors have observed, the field of law enforcement is inherently sensitive, as it concerns “the very heart of the relation between individuals and public authorities”. G. GONZÁLEZ FUSTER, *Artificial Intelligence and Law Enforcement– Impact on Fundamental Rights*, Policy Department for Citizens’ Rights and Constitutional Affairs – Directorate-General for Internal Policies, 2020, 67. Available at: [https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU\(2020\)656295](https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2020)656295). Similarly “This is a context where the stakes are high. Given the unique competencies that law enforcement agencies have, the relationship between these agencies and the individuals and communities they serve is often one of power imbalance”. UNICRI, INTERPOL, *Responsible AI Innovation in law enforcement. AI toolkit. 01. Introduction to Responsible AI Innovation*, cited, 15.

risks arise, was incorporated into the AI Act³⁰⁹. It was introduced at a later stage of the legislative process, following the amendments proposed by the European Parliament³¹⁰. While the final version of the AI Act ultimately endorsed the Parliament’s proposal, it significantly narrowed its scope of application, limiting the obligation to “*bodies governed by public law, private entities providing public services, and deployers of high-risk AI systems referred to in points 5(b) and (c) of Annex III*”³¹¹. This limitation remains open to debate, as it arguably excludes other actors whose activities may equally affect fundamental rights³¹². For the purposes of the present study, however, it is submitted that law enforcement authorities, together with private entities acting on their behalf, fall within the personal scope of this obligation.

In conducting such assessments, particular attention must be directed to the principle of non-discrimination, given that AI system deployment and use may profoundly affect equality and fairness, especially within the sensitive law enforcement context. Law enforcement authorities are accordingly under an obligation to conduct a thorough assessment of the potential impact of AI systems on distinct social groups, with a view to ensuring that their operation does not give rise to discriminatory outcomes. This entails the establishment of robust procedural frameworks for testing and evaluating the diversity and representativeness of the datasets employed, including with respect to specific social groups or particular use cases, as well as a rigorous examination of the underlying algorithmic logic for analogous concerns. Moreover, law enforcement authorities are required to implement systematic processes capable of identifying whether certain groups are disproportionately affected by the deployment of such systems, and to ensure that adequate mechanisms are in place to detect, mitigate, and remedy instances of bias, discrimination, or suboptimal performance.

Specifically, law enforcement authorities who act as deployers must perform an assessment that includes³¹³: (a) a description of the process in which the high-risk AI system will be used, consistent with its intended purpose; (b) a description of the period and frequency of use of each high-risk AI

³⁰⁹ For a discussion on FRIA in the AI Act see DONET, *The Fundamental Rights Impact Assessment By Deployers Of Artificial Intelligence Systems In The Regulation*, cited, 523-566; MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act*, cited.

³¹⁰ While the Commission’s proposal emphasised a human-centric approach and the protection of fundamental rights, these principles were not meaningfully operationalised within the risk-based framework of the draft legislation. In response, the European Parliament introduced a FRIA aimed at evaluating the potential implications of AI systems for fundamental rights. Nevertheless, the final version of the AI Act substantially narrowed the FRIA’s content, scope, and participatory mechanisms as initially envisaged in the Parliament’s amendments.

³¹¹ AI Act, Article 27.

³¹² DONET, *The Fundamental Rights Impact Assessment By Deployers Of Artificial Intelligence Systems In The Regulation*, cited, 538.

³¹³ AI Act, Article 27(5) specifies that “*The AI Office shall develop a template for a questionnaire, including through an automated tool, to facilitate deployers in complying with their obligations under this Article in a simplified manner*”. While the FRIA does not provide a detailed outline of its required content, this absence should not be understood as diminishing the rigor or depth of the assessment to be conducted. MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act*, cited, 8.

system; (c) the identification of categories of natural persons and groups likely to be affected by its use in the specific context; (d) the identification and assessment of the specific risks that may impact the identified categories or groups, taking into account the information provided by the system's provider pursuant to Article 13; (e) a description of the implementation of human oversight measures, in accordance with the system's instructions for use; and (f) the measures to be taken in the event that such risks materialise, including arrangements for internal governance and complaint-handling mechanisms.

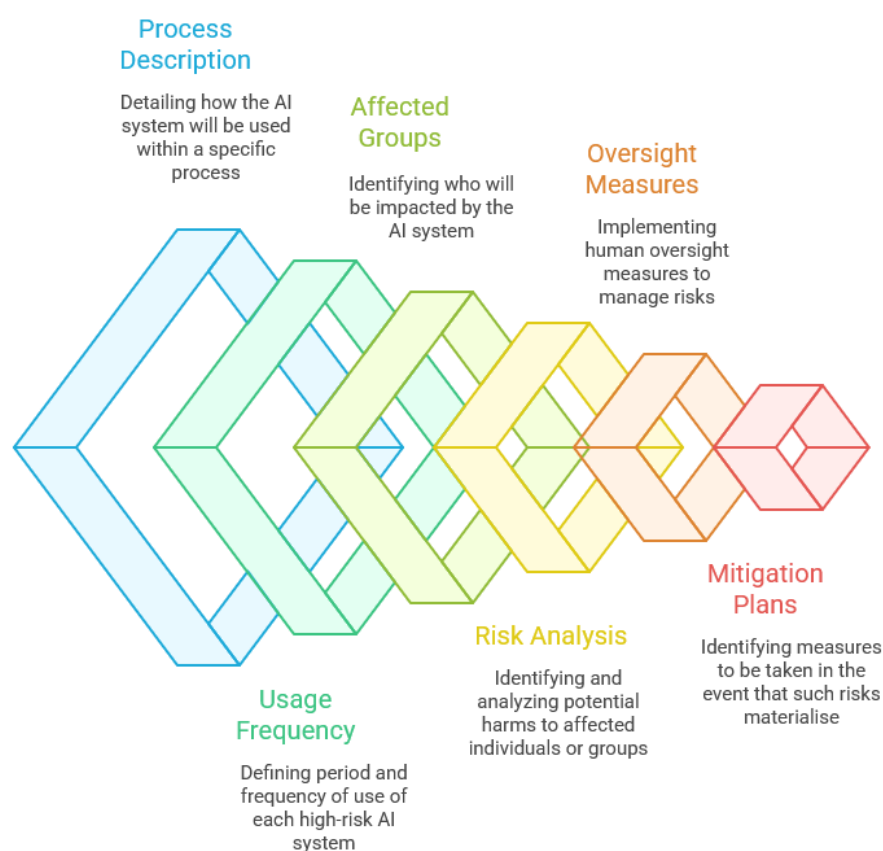


Figure 13: Article 27 - Content of the Fundamental Rights Impact Assessment. Own elaboration with Napkin.

The contextual dimension of AI systems is of fundamental importance. Understanding how fundamental rights are contextualised requires an examination of the legal frameworks and judicial interpretations that define their scope and meaning, as well as the distinct nuances that arise across national jurisdictions. Equally, the context in which an AI system is deployed is relevant, as it shapes the system's interaction with the broader socio-technical environment in which it functions. Therefore, law enforcement authorities should carefully consider the conditions under which an AI system is intended to operate, in order to fully understand the potential risks and benefits associated with its use. In addition, Article 27(c) and (d) of the AI Act are particularly significant, as they enable

an assessment of whether and how an AI system impacts different groups, thereby informing the adoption of appropriate mitigating and protective measures.

An illustrative example of a FRIA conducted pursuant to Article 27 of the AI Act in the specific context of law enforcement is the model developed within the already mentioned ALIGNER project³¹⁴, which integrates both fundamental rights and AI ethics considerations³¹⁵.

The FRIA, also in the context of law enforcement, is intended to complement, and not replace, other assessments mandated by law. For instance, Article 27 of the LED requires to carry out a Data Protection Impact Assessment (hereinafter “DPIA”) where processing activities, particularly those involving new technologies, are likely to result in a high risk to the rights and freedoms of natural persons, considering the nature, scope, context, and purposes of the processing³¹⁶. As Alessandro Mantelero notes, the FRIA exhibits the same core features as the DPIA, including: (i) an *ex ante* approach³¹⁷, (ii) a rights-based focus on risk assessment³¹⁸, (iii) a circular iterative structure that

³¹⁴ Available at: <https://aligner-h2020.eu/fundamental-rights-impact-assessment-fria/>. Apart from the FRIA developed within the ALIGNER project, several other tools have been designed to assess the potential impact of AI systems on fundamental rights, reflecting a growing effort to operationalise ethical and legal principles in AI governance. Notably, INTERPOL and UNICRI have developed the Toolkit for Responsible AI Innovation in Law Enforcement (2024), which includes a Risk Assessment Questionnaire aimed at assisting law enforcement authorities (LEAs) in identifying and evaluating potential risks in relation to four core ethical principles. UNICRI, INTERPOL, *Responsible AI Innovation in law enforcement. AI toolkit – 05. Risk Assessment Questionnaire*, cited. A slightly different instrument in terms of purpose and scope is the Accountability Principles for Artificial Intelligence (AP4AI) in the Internal Security Domain which outlines a framework for accountability and governance of AI systems in internal security contexts. B. AKHGAR, P.S. BAYERL, K. BAILEY, R. DENNIS, H. GIBSON, S. HEYES, A. LYLE, A. RAVEN, F. SAMPSON, *Accountability Principles for Artificial Intelligence (AP4AI) in the Internal Security Domain*, February 2022. Available at: https://www.europol.europa.eu/cms/sites/default/files/documents/Accountability_Principles_for_Artificial_Intelligence_AP4AI_in_the_Internal_Security_Domain.pdf.

³¹⁵ D. CASABURO, I. MARSH, *Ensuring fundamental rights compliance and trustworthiness of law enforcement AI systems: the ALIGNER Fundamental Rights Impact Assessment*, in *AI Ethics*, 4, 2024, 1569–1582. The ALIGNER Fundamental Rights Impact Assessment (AFRIA) represents an important effort to ensure that AI systems used in law enforcement comply with fundamental rights and uphold public trust. Nevertheless, as Mantelero observes, the limited experience with Human Rights Impact Assessments (HRIAs) in the context of AI, combined with the scarcity of methodological debate in this area, indicates the need for a more in-depth analysis of possible tools and approaches for conducting FRIAs tailored to AI applications.

³¹⁶ It shall be conducted prior to the processing and shall contain at least a general description of the envisaged processing operations, an assessment of the risks to the rights and freedoms of data subjects, the measures envisaged to address those risks, safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Directive, taking into account the rights and legitimate interests of the data subjects and other persons concerned. See MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act*, cited.

³¹⁷ Both the Data Protection Impact Assessment (DPIA) and the Fundamental Rights Impact Assessment (FRIA) are required to be carried out in advance, respectively, before the processing of personal data and before the deployment of a high-risk AI system.

³¹⁸ The risk assessment should be conducted with rights as its focal point: while the DPIA concentrates on the right to the protection of personal data, the FRIA extends the analysis to a broader set of fundamental rights potentially affected by the deployment of AI systems.

follows the product/service throughout its lifecycle³¹⁹ and (iv) an expert-based nature³²⁰. However, these assessments serve distinct yet complementary objectives: while the DPIA focuses specifically on identifying and evaluating risks to the right to data protection, the FRIA is concerned with assessing broader risks that may affect the enjoyment of a wider spectrum of fundamental rights. Consequently, in all situations falling within the scope of Article 27, both assessments should be carried out by law enforcement authorities. Similar to the DPIA, the FRIA should be performed in relation to a single AI system deployed by law enforcement authorities for a specific law enforcement purpose³²¹. The assessment should consider the system's predefined purposes and conditions of use, including trigger mechanisms, duration and frequency of operation, affected individuals or groups, and the relevant geographical areas. The FRIA and DPIA shall be carried out jointly, if any of the obligations established for FRIA are already being met by the DPIA³²², or in coordination.

A FRIA is best understood as a process composed of several distinct phases, following the “Plan-Do-Check-Act cycle”, which ensures a systematic and iterative approach³²³. The initial step requires determining whether a FRIA is mandated under Article 27(1), which delimits the assessment's scope. This necessitates that law enforcement authorities first establish – excluding cases where the AI system's use is prohibited under Article 5 – whether the AI system in question falls within the high-risk category. Once the need for a FRIA has been established, it is essential to gather the necessary information to define the context, plan the process, and develop a detailed understanding of the AI system and its intended use (so-called “Planning and scoping phase”). These preparatory phases make it possible to identify the known and reasonably foreseeable risks that the AI system may pose to fundamental rights³²⁴. The subsequent phase involves the analysis of potential risk scenarios, which requires an assessment of risks through a risk matrix (so-called “risk-analysis phase”), which requires considering both the impact of risk, that is, the degree of severity of the potential harm or prejudice

³¹⁹ The circular and iterative nature of risk assessment underscores the need for continuous evaluation throughout the system's lifecycle. Owing to the intrinsic link between potential risks and system design, it is advisable to initiate the assessment at the earliest stage of defining the deployer's strategy for the intended use of an AI system. The process should subsequently be revisited whenever substantial changes occur in the system's design, operation, or deployment context. MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act*, cited, 8. AI Act, Recital 96 “*The impact assessment should be performed prior to deploying the high-risk AI system, and should be updated when the deployer considers that any of the relevant factors have changed*”.

³²⁰ MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act*, cited, 4. The participation of experts in the FRIA process is of paramount importance, as it contributes to the overall quality and reliability of the assessment while also addressing typical limitations inherent in questionnaire-based methodologies, including unintentional bias in question design, respondent confirmation bias, and incomplete or inconsistent information. *Ibidem*, 12.

³²¹ The FRIA must be carried out prior to the initial deployment of a high-risk AI system. The deployer may rely on a previously conducted FRIA or on an existing assessment undertaken by the provider.

³²² AI Act, Article 27(4).

³²³ DONET, *The Fundamental Rights Impact Assessment By Deployers Of Artificial Intelligence Systems In The Regulation*, cited, 550.

³²⁴ “*The materialization of these risks referred to in Article 27(1)(f) is distinct from the materialisation of harm and ex post remedies*”. MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act*, cited, 9.

experienced by the affected individuals, and the likelihood of the risk, meaning the probability of its occurrence.

Once the initial risk has been assessed, appropriate mitigation measures must be put in place, a process known as the “risk management phase”. Article 9 of the AI Act is particularly instructive in this regard, as it establishes the framework governing continuous risk management and mitigation³²⁵. While both start from the same situation of “high-risk”³²⁶, the FRIA has a narrower scope than the risk management system. More specifically, Article 9 mandates the AI provider to implement a comprehensive and ongoing risk management system, defined as an iterative process to identify and analyse, estimate and evaluate, risks associated with high-risk AI systems during their entire lifecycle: from their initial design and development, through deployment, and into the post-market monitoring phase, where new risks may emerge³²⁷. Providers shall define and adopt clear policies, procedures, and lines of responsibility for risk management, as well as to conduct a risk assessment in order to understand which measures should be put in place to mitigate risks and how strong those measures need to be³²⁸. The ultimate objective of the risk management system is either to eliminate such risks altogether or, where elimination is not feasible, to reduce them to a level that can be deemed acceptable in light of the system’s intended purpose and the fundamental rights potentially affected by its use³²⁹.

As the management and mitigation of risks are common objectives of both the risk management system and the FRIA, the measures set out in Article 9(5) may also be relevant in the context of

³²⁵ For a discussion on the differences between impact assessment and risk management in data protection see E.C. DONET, *The Fundamental Rights Impact Assessment By Deployers Of Artificial Intelligence Systems In The Regulation*, cited, 527. For a discussion on the risk management framework see P.S. CASTELLANO, *Risk management systems as a specific obligation for high-risk artificial intelligence systems in article 9 of the regulation*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 568-594. Reference may also be made to ISO 31010:2010, which provides detailed methodologies for conducting risk assessments. In the field of information security management, a pertinent example is UNE-EN ISO/IEC 27001:2023, the European standard transposing the international standard ISO/IEC 27001:2022 on information security management system requirements. This framework is further elaborated by ISO/IEC 27005, which sets out guidance for implementing a process-oriented risk management approach, thereby facilitating compliance with the security risk management obligations under ISO/IEC 27001.

³²⁶ They both represent obligations of providers and deployers of high-risk AI systems.

³²⁷ On the AI Act’s risk management system, see J. SCHUETT, *Risk Management in the Artificial Intelligence Act*, in *European Journal of Risk Regulation*, 2024, 15, 367-385.

³²⁸ “Assessing risks starts with the identification of hazards that stem from the use, or misuse, of the product, taking into consideration all types of harms that may be involved (such as physical harms, breaches of privacy, or discrimination). Then assessors determine likelihood and severity levels for each hazardous situation, using guidance and categories from a risk management standard”. H. POUGET, R. ZUHDI, *AI and Product Safety Standards Under the EU AI Act*, 5 March 2024. Available at: <https://carnegieendowment.org/research/2024/03/ai-and-product-safety-standards-under-the-eu-ai-act?lang=en>.

³²⁹ “Acceptability means that the risk may remain high, albeit mitigated if justified by the potential benefits associated with AI”. MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act*, cited, 4-5. On the acceptability of risk see H. FRASER, J.-M. BELLO Y VILLARINO, *Acceptable Risks in Europe’s Proposed AI Act: Reasonableness and Other Principles for Deciding How Much Risk Management Is Enough*, in *European Journal of Risk Regulation*, 2023, 2-3. J. LAUX, S. WACHTER, B. MITTELSTADT, *Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk*, in *Regulation & Governance*, 18, 2024, 3-32.

Article 27, although they are more technical in nature. Moreover, Article 27 specifically refers to human oversight measures, internal governance and complaint mechanisms³³⁰.

As a result of the risk management process, the residual risk is established, meaning the level of risk that persists following the implementation of control measures designed to mitigate or reduce exposure to the identified risk factors. Therefore, risk management measures are designed to ensure that the residual risks associated with each identified hazard, as well as the overall residual risk of the high-risk AI system, are reduced at a level considered acceptable³³¹. Accordingly, the objective of the FRIA is to ensure that, given the initial risk, the application of appropriate measures or controls reduces the residual risk to a level below the threshold of acceptability.

To mitigate the administrative burden on deployers, the AI Office is entrusted with developing a standardised template³³², and potentially an automated tool, to support compliance with the FRIA obligation³³³. Nonetheless, from my perspective, the full automation of the FRIA is not advisable, as it could undermine the assessment's essential focus on the broader socio-technical dynamics surrounding the deployment and use of a high-risk AI system for its specific purpose. Furthermore, sector-specific factors – such as policing practices, demographic composition, or situational constraints – must be duly integrated into the assessment framework³³⁴, given their significant bearing on both the system's impact and its associated risks.

It should be emphasised that the FRIA is to be conducted *after* the provider has completed the risk assessment and implemented the risk management system in accordance with Article 9 of the AI Act. More specifically, while the risk management system established under Article 9 provides a general framework for assessing the risks of high-risk AI systems, the FRIA places emphasis on the deployer's role in managing context-specific risks associated with the particular use of the AI system³³⁵. For example, in the case of a RBI system used for law enforcement, Article 9 would govern the general assessment of risks in the system's design, whereas the FRIA would require the law enforcement authority acting as deployer to evaluate how those risks manifest in a specific deployment context – for instance, during real-time identification in public spaces – and to implement

³³⁰ AI Act, Article 27(1)(e) and (f).

³³¹ The literature has identified the following measures: 1) mitigate or reduce the initial risk by implementing controls that reduce the risk below the threshold defined as acceptable, by reducing the impact caused by a risk scenario or the likelihood of a risk scenario materialising; 2) avoid or eliminate the risk by nullifying, excluding or replacing the design element or functionality, even if this is not always feasible as it can sometimes result in the loss of essential functionality; 3) according to “general risk management theory” is to ignore or assume the risk.

³³² The AI Office shall develop a template for a questionnaire. For a critical discussion on such questionnaires see MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act*, cited, 9: “a purely questionnaire-based approach, and even worse its automation (which requires a high degree of uniformity), cannot fully capture the contextual nature of the FRIA and needs to be integrated with a methodology for risk quantification and management”.

³³³ AI Act, Article 27(5).

³³⁴ MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act*, cited, 12.

³³⁵ *Ibidem*, 7.

appropriate mitigation measures accordingly. Unlike a purely horizontal approach, which provides broad, cross-sectoral rules but may overlook sector-specific challenges, the FRIA provided for by the AI Act ensures that concrete risks are assessed in relation to the intended use of the AI system³³⁶. The redundancy mechanism inherent in the AI Act, insofar as risk management is concerned, underscores the importance of ensuring compliance and safeguarding fundamental rights in the use of AI systems, particularly in highly sensitive sectors and by actors that hold a significant imbalance of power over the individuals concerned.

The introduction of the FRIA thus represents an overall positive development. Indeed, risks linked to individual vulnerabilities and context-specific biases often lie beyond what a provider can foresee or manage at a general level under Article 9. However, its effectiveness presupposes that it extends beyond a descriptive enumeration of potentially affected rights accompanied by generic mitigation proposals. Rather, it requires a demonstrable evidentiary link between the identified adverse impacts of AI systems on fundamental rights and the adequacy, proportionality, and effectiveness of the measures adopted to mitigate or eliminate them. Accordingly, the effectiveness of the FRIA presupposes that deployers either possess, or have access to, the necessary expertise to conduct the assessment from a sound methodological perspective³³⁷, and possess a sufficient understanding of the AI system's functioning. To this end, the transparency requirements and obligations outlined above become crucial enablers of an effective FRIA, as does the promotion of AI literacy among deployers³³⁸.

Requiring deployers to notify the market surveillance authority of their FRIA results may further enhance accountability and ensure the effectiveness of the assessment process³³⁹. Nevertheless,

³³⁶ Despite the positive introduction of the FRIA, the final version of the AI Act significantly narrowed the scope of the instrument as originally proposed in the European Parliament's amendments. First, the obligation to conduct a FRIA now applies only to a limited subset of deployers. Second, the participatory measures initially envisaged by the Parliament have been removed. In particular, the Parliament's text had required deployers, to the greatest extent possible, to involve representatives of individuals or groups likely to be affected by the high-risk AI system (such as relevant stakeholders as well as representatives of groups of persons likely to be affected by the AI system) in order to collect relevant information for the assessment. The importance of participatory measures in ensuring legitimacy and accountability in impact assessment processes has been emphasised by A. MANTELERO, *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*, Springer, 2022.

³³⁷ As discussed above, the FRIA – like the DPIA and other forms of impact assessment – is inherently expert-based in nature. The complexity of conducting a FRIA is clearly demonstrated in A. MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act*, cited. In relation to the DPIA, see A. MANTELERO, *Comment to Articles 35 and 36*, in M. COLE, F. BOEHM (Eds.), *GDPR Commentary*, Edward Elgar Publishing, 2019. Examples of how impact assessments have been implemented, and from which valuable insights can be drawn, include the *Algorithmic Impact Assessment Tool* developed by the Government of Canada, which provides a risk index for algorithmic impacts; the *Fundamental Rights and Algorithms Impact Assessment (FRAI)* proposed by the Dutch Ministry of the Interior and Kingdom Relations; the *HUDERIA* (Human Rights, Democracy, and the Rule of Law Impact Assessment) designed by the Alan Turing Institute within the framework of the Council of Europe's Ad hoc Committee on Artificial Intelligence and UNESCO's *Ethical Impact Assessment for Artificial Intelligence*. These initiatives illustrate the increasing institutional effort to develop structured, expert-driven methodologies for assessing the ethical, social, and legal implications of AI systems, thereby contributing to the broader objective of ensuring protection of fundamental rights.

³³⁸ AI Act, Article 4.

³³⁹ AI Act, Article 27(3).

Article 46(1) provides an exemption from this obligation in exceptional cases, where overriding concerns – such as public security, the protection of human life or health, environmental safety, or the integrity of critical infrastructure – justify withholding such notification.

A critical limitation of Article 27 lies in its circumscribed scope of application, which restricts the FRIA obligation exclusively to deployers of high-risk AI systems³⁴⁰. As noted above, this determination requires law enforcement authorities to establish whether the AI system in question falls within the high-risk category. In this author's contention, however, this restriction fails to account for the contextual nature of fundamental rights impacts: even AI systems not classified as high-risk *ex ante* may generate significant fundamental rights implications when deployed in sensitive operational environments.

This concern assumes particular relevance in the law enforcement context. Given the inherent power asymmetries characterizing interactions between law enforcement authorities and individuals, the potential for rights infringement exists regardless of whether the deployed AI system formally meets the high-risk threshold. Law enforcement activities routinely implicate fundamental rights including liberty, privacy, fair trial guarantees, non-discrimination, and human dignity. The deployment of any AI system – whether classified as high-risk or otherwise – within such a sensitive domain introduces variables that may affect these rights in ways not fully captured by the AI Act's risk classification framework, which focuses primarily on the technical characteristics and intended purpose of AI systems rather than on the institutional context of their deployment.

In this author's view, the FRIA requirement should therefore extend to all AI systems deployed in fundamentally rights-sensitive sectors, irrespective of their formal risk classification. Such an expansion would bring the regulatory framework into closer alignment with the principle that fundamental rights assessments ought to be triggered not solely by system characteristics, but by the confluence of system capabilities and deployment context. In the law enforcement setting specifically, a universal FRIA requirement would give formal recognition to the fact that the sector's intrinsic connection to coercive state power and fundamental rights restrictions demands heightened scrutiny of all AI deployments, ensuring that even ostensibly lower-risk systems are subjected to rigorous evaluation before being operationalised.

3.2.From data to discrimination: how the AI Act confronts bias in AI-powered law enforcement

³⁴⁰ AI Act, Recital 93 “Deployers are best placed to understand how the high-risk AI system will be used concretely and can therefore identify potential significant risks that were not foreseen in the development phase, due to a more precise knowledge of the context of use, the persons or groups of persons likely to be affected, including vulnerable groups”.

The AI Act also seeks to prevent biased outcomes and discrimination through provisions on data and data governance³⁴¹, recognising the vital importance of high-quality data for the proper functioning of AI systems³⁴². Bias in data can emerge from numerous sources. First and foremost, the historical data used to train AI models can perpetuate long-standing societal biases by reflecting past prejudices and discriminatory practices³⁴³. For example, if a neighbourhood has historically been subject to disproportionate policing as a result of racial or socio-economic prejudices, an AI system trained on this data may identify it as being at a higher risk of criminal activity. As highlighted in Chapter 2, such dynamics can give rise to a self-reinforcing feedback loop, whereby biased data lead to biased predictions, which in turn perpetuate and legitimise existing structural inequalities³⁴⁴.

Biased outcomes may also derive from representational bias³⁴⁵, which occurs when the data used to train an AI system does not adequately reflect the diversity of the population in which the system is intended to be deployed³⁴⁶. If certain demographic groups – such as those defined by gender, ethnicity, age, or disability – are underrepresented or misrepresented in the training data, the system’s outputs may systematically disadvantage those groups. This form of bias is particularly concerning in high-risk contexts such as in relation to biometric identification, where skewed datasets can lead to higher error rates or misidentification for specific segments of the population. Article 10 on data governance is the AI Act’s most explicit anti-discrimination provision, mandating the “*examination*

³⁴¹ EUROPOL, *AI and policing The benefits and challenges of artificial intelligence for law enforcement*, cited, 32-33. In simplified terms, “data are units of information describing quantity, quality, facts, statistics or other basic units of meaning, collected for reference or analysis”. OECD, OECD Glossary of Statistical Terms, 2008. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2008/09/oecd-glossary-of-statistical-terms_g1gh9ad7/9789264055087-en.pdf

³⁴² The absence of adequate and high-quality data may have far-reaching consequences, as it shapes the functioning and outputs of an AI system from the very design phase. Insufficient or flawed data not only compromise the validity of the resulting AI solution but may also pose significant risks to individuals’ security and the protection of their fundamental rights. Accordingly, the quality of an AI system’s results is directly contingent upon the quality, accuracy, and representativeness of the data on which it is developed and trained. M.L. CORERA, *Data And Data Governance And Connections To Data Protection Principles In Article 10 Of The Artificial Intelligence Act*, in in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 595.

³⁴³ R. RICHARDSON, J.M. SCHULTZ, K. CRAWFORD, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, in *New York University Law Review Online*, 94, 2019, 15-55. “In a context where discrimination has historically been an issue, there is a risk that historical law enforcement data will reflect the individual and institutional prejudices that have had a disproportionate impact on certain individuals and groups”. UNICRI, INTERPOL, *Responsible AI Innovation in law enforcement. AI toolkit*, 02. *Principles for Responsible AI Innovation*, cited, 24.

³⁴⁴ C. VEAL, M. RAPER, P. WATERS, *The perils of feedback loops in machine learning: predictive policing*, published on Gilbert+Tobin, 20 February 2023. Available at: <https://www.gtlaw.com.au/insights/the-perils-of-feedback-loops-in-machine-learning-predictive-policing>. XENIDIS, SENDEN, *EU Non-discrimination Law in the Era of Artificial Intelligence*, cited, 9.

³⁴⁵ EUROPOL, *AI and policing The benefits and challenges of artificial intelligence for law enforcement*, cited, 32. N. SHAHBAZI, Y. LIN, A. ASUDEH, H.V. JAGADISH, *Representation Bias in Data: A Survey on Identification and Resolution Techniques*, in *ACM Computing Surveys*, 55(13), 2023, Article No. 293, 1-39. Available at <https://doi.org/10.1145/3588433>.

³⁴⁶ One of the main issues in data used to train algorithms is that such data are generally collected from the web, social media or are synthetic data, mostly originated in North America and Europe, where white-skinned men are prevalent. K. CRAWFORD, *Artificial intelligence’s White Guy Problem*, published on the *New York Times*, 25 June 2016. Available at <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>.

in view of possible biases that are likely to affect the health and safety of persons, have a negative impact on fundamental rights or lead to discrimination prohibited under Union law” (Article 10(2)(f)), along with “appropriate measures to detect, prevent and mitigate” such biases (Article 10(2)(g)).

To achieve this objective, the datasets used to train, validate, test, and operationalise AI systems must meet rigorous quality standards, underpinned by a governance framework that goes beyond mere data control and management³⁴⁷. Data governance³⁴⁸ specifically refers to a set of rules, processes, and responsibilities that determine how data is collected, managed, accessed, and used throughout its lifecycle with the goal of ensuring that data is accurate, reliable, secure, and ethically handled.

Although Article 10 of the AI Act does not explicitly refer to good data governance, it implicitly incorporates its principles, requiring that training, validation, and testing datasets comply with the quality criteria set out in paragraphs 2 to 5 of that provision³⁴⁹. Specifically, Article 10(2) sets out the measures required to ensure sound data governance and management practices, so that data are fit for the AI system’s intended purpose³⁵⁰.

³⁴⁷ In recent years, the European Union (EU) has increasingly recognised data as a strategic asset for innovation, economic growth and social welfare. Within the framework of the European Commission’s priority “*A Europe Fit for the Digital Age*” (2019–2024), the *European Strategy for Data* (2020) seeks to establish a single European data space, a common market for data. Two legislative instruments, notably the *Data Governance Act* (2022) and the *Data Act* (2023), operationalize this vision by defining governance mechanisms and fair access conditions for both personal and non-personal data. These initiatives are further embedded in the *Digital Compass 2030*, which outlines Europe’s “*Digital Decade*” targets across four pillars: digital skills, secure and sustainable digital infrastructure, the digital transformation of businesses, and the digitalisation of public services.

³⁴⁸ The Data Governance Institute has defined “data governance” as “*the exercise of decision-making and authority in data-related matters*”, and more broadly, as “*a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods*”. Available at: <https://datagovernance.com/the-data-governance-basics/definitions-of-data-governance/>. No explicit definition of data governance is provided in either the AI Act or the ISO/IEC 22989:2022 (Information Technology — Artificial Intelligence — Concepts and Terminology) standard. Nonetheless, several scholars have attempted to conceptualize data governance within the broader frameworks of AI regulation and information management. There appears to be a general consensus in associating data governance with three main ideas: (1) the recognition of data as a valuable organisational asset that must be effectively managed; (2) the establishment of clear responsibilities in decision-making processes, including rights and corresponding duties; and (3) the formulation of guidelines and standards to ensure data quality, integrity, and appropriate use. M. SALVADOR SERNA, *Inteligencia artificial y gobernanza de datos en la Administración Pública: sentando las bases para su integración a nivel corporativo*, in *Repensando la administración pública: administración digital e innovación pública*, INAP, 2021, 126–148. To these characteristics, some authors add a fourth: the need for strategic leadership from management in the establishment of a data governance framework. Such leadership is essential to ensure that data governance aligns coherently with the organisation’s objectives, culture, and, importantly, the applicable regulatory requirements. M.L. CORERA, *Data And Data Governance And Connections To Data Protection Principles In Article 10 Of The Artificial Intelligence Act*, cited, 598.

³⁴⁹ By contrast, when the development of high-risk AI systems does not require the training of AI models, paragraphs 2 to 5 apply only to the testing data sets.

³⁵⁰ AI Act, Article 10(2). Those practices shall concern in particular: (a) the relevant design choices; (b) data collection processes and the origin of data, and in the case of personal data, the original purpose of the data collection; (c) relevant data-preparation processing operations, such as annotation, labelling, cleaning, updating, enrichment and aggregation; (d) the formulation of assumptions, in particular with respect to the information that the data are supposed to measure and represent; (e) an assessment of the availability, quantity and suitability of the data sets that are needed; (f) examination in view of possible biases that are likely to affect the health and safety of persons, have a negative impact on fundamental rights or lead to discrimination prohibited under Union law, especially where data outputs influence inputs for future

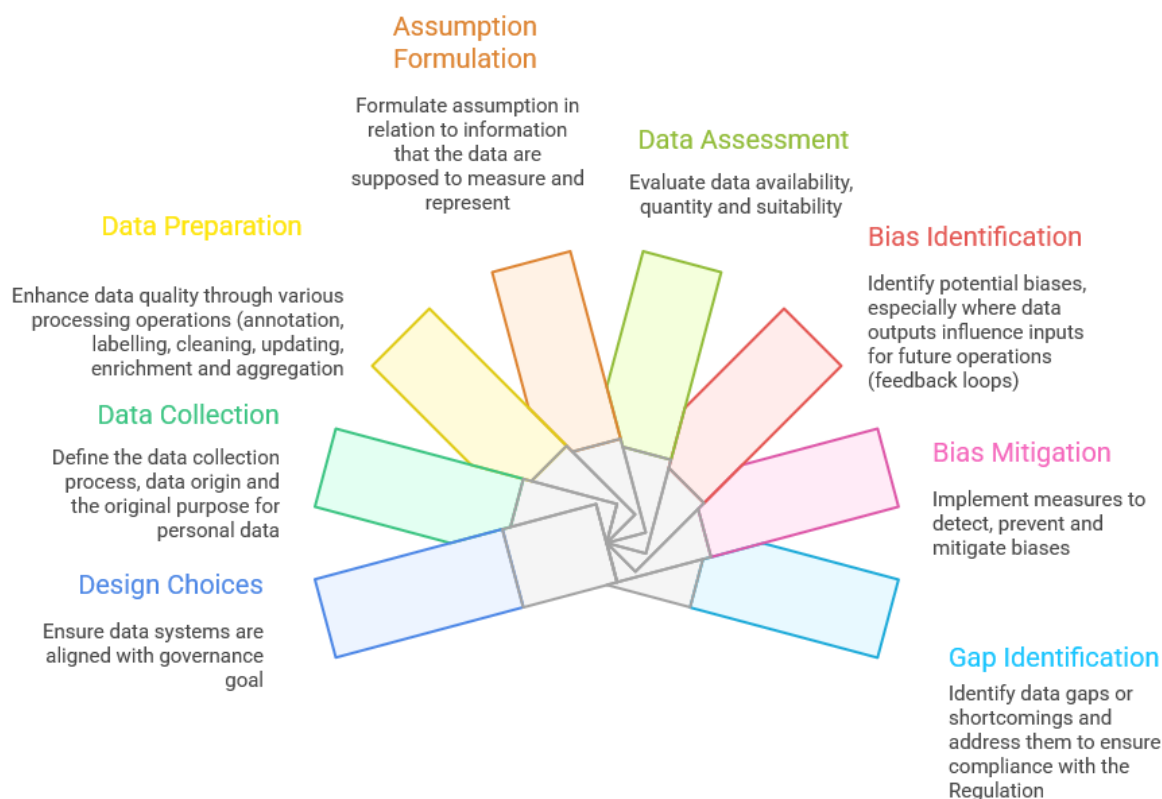


Figure 14: Data governance and management practices under Article 10(2). Own elaboration with Napkin.

The notion of data quality refers to the fact that data must be “*fit for use*”³⁵¹, that is, suitable and reliable for the specific objectives for which they are collected and processed. The principle of high data quality had already found its way into the European legal framework with the entry into force of the GDPR, albeit in a slightly different form. Article 5 of the GDPR – and Article 7 of the LED in a

operations; (g) appropriate measures to detect, prevent and mitigate possible biases identified according to point (f); (h) the identification of relevant data gaps or shortcomings that prevent compliance with this Regulation, and how those gaps and shortcomings can be addressed. When personal data are processed, data governance and management practices should include, in the case of personal data, transparency about the original purpose of the data collection. These practices must, in particular, form part of the quality management system. AI Act, Article 17(1)(f). The quality management system shall include “*f) systems and procedures for data management, including data acquisition, data collection, data analysis, data labelling, data storage, data filtration, data mining, data aggregation, data retention and any other operation regarding the data that is performed before and for the purpose of the placing on the market or the putting into service of high-risk AI systems*”. By contrast, no such reference appears in relation to the risk management system under Article 9 of the AI Act. Similarly, data governance practices are explicitly mentioned only in the context of the technical documentation requirements set out in Annex IV: “*(d) where relevant, the data requirements in terms of datasheets describing the training methodologies and techniques and the training data sets used, including a general description of these data sets, information about their provenance, scope and main characteristics; how the data was obtained and selected; labelling procedures (e.g. for supervised learning), data cleaning methodologies (e.g. outliers detection)*”.

³⁵¹ R.Y. WANG, D.M. STRONG, *Beyond Accuracy: What Data Quality Means to Data Consumers* in *Journal of Management Information Systems*, 12(4), 1996, 6.

slightly different way³⁵² – introduced the principles of data accuracy and updating³⁵³, as well as the principle of data minimization³⁵⁴, which can be seen as precursors to the principle of high-quality data established for AI applications under Article 10 of the AI Act.

As the literature has pointed out, assessing data quality for AI systems primarily involves considerations of completeness, accuracy, precision, consistency, timeliness, reliability, and correctness³⁵⁵. A dataset must, first and foremost, be complete, meaning that any missing data should be identified and appropriately addressed. This may involve reconstructing incomplete datasets through integration with other datasets, collecting additional data from supplementary sources, or, where data are irreparably compromised, excluding the affected datasets altogether. Data must also be accurate and precise, aligning with the intended purpose of the AI system. Assessing precision requires comparison with established reference values, an exercise that may be straightforward in some cases (for example, binary gender data) but considerably more complex in others. Furthermore, data should exhibit internal consistency and be properly harmonized when integrated across multiple datasets. Given the rapid pace at which data are generated and collected, timeliness is also critical: outdated data risk producing misleading or irrelevant outcomes. Reliability demands that data be genuine and derived from credible sources, while correctness requires vigilance against errors introduced through human input or other manual processes.

In line with the above discussion, Article 10(3) elaborates concrete data quality criteria, requiring that training, validation, and testing datasets be relevant, sufficiently representative, and “*to the best extent*

³⁵² LED, Article 7(2) “*Member States shall provide for the competent authorities to take all reasonable steps to ensure that personal data which are inaccurate, incomplete or no longer up to date are not transmitted or made available. To that end, each competent authority shall, as far as practicable, verify the quality of personal data before they are transmitted or made available. As far as possible, in all transmissions of personal data, necessary information enabling the receiving competent authority to assess the degree of accuracy, completeness and reliability of personal data, and the extent to which they are up to date shall be added*”.

³⁵³ The data controller must remain constantly aware of the purposes underlying the processing activity. Without reference to these purposes, it is impossible to assess whether the data or dataset being used is appropriate, accurate, and up to date. See F. PIZZETTI, *La protezione dei dati personali e la sfida dell’Intelligenza Artificiale*, in F. PIZZETTI (Ed.), *Intelligenza Artificiale, protezione dei dati personali e regolazione*, Giappichelli, 2018, 61.

³⁵⁴ Pursuant to the principle of data minimisation enshrined in Article 5(1)(c) GDPR, personal data must be adequate, relevant, and limited to what is necessary in relation to the purposes of the processing. This principle requires controllers to reduce the quantity of data processed to what is strictly necessary to achieve the intended purposes. As Giuseppe Mobilio notes, data minimisation is connected not only to the purpose of collection and the categories and quantity of data processed but also to the temporal dimension of their use, thereby engaging the principle of storage limitation G. MOBILIO, *L’intelligenza artificiale e le regole giuridiche alla prova: il caso paradigmatico del GDPR*, in *Federalismi.it*, 16, 2020, 279.

³⁵⁵ The metrics used to determine data quality are not uniform, and various authors propose different lists of relevant criteria. See, among others, A. BURT, *Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning Models*, in *The Future of Privacy Forum*, 2018, 4; L. CAI, Y. ZHU, *The Challenges of Data Quality and Data Quality Assessment in the Big Data Era*, in *Data Science Journal*, 14, 2015, 5–6. Franco Pizzetti warns against what he terms “data trawling” referring to the indiscriminate collection of data without due consideration of its purpose, accuracy, or integrity. PIZZETTI, *La protezione dei dati personali e la sfida dell’Intelligenza Artificiale*, cited, 62.

*possible*³⁵⁶, accurate, complete, and error-free in view of the system's intended use. Moreover, these datasets must demonstrate appropriate statistical characteristics, including, where relevant, with regard to the persons or groups of persons affected by the deployment of the high-risk AI system.

These provisions show the recognised vital importance of careful data preparation for the correct functioning of the AI system and are primarily directed at the providers of high-risk AI systems, mandating them to consider the specific context and purpose for which the AI system is to be deployed.

This, in turn, presupposes continuous cooperation with the deployer, who is in the most appropriate position to ascertain and define the intended use of the AI system. In Chapter 1, I briefly outlined the AI lifecycle as a sequential yet iterative process, beginning with the definition of the problem for which the AI system is intended and culminating in its deployment and monitoring. This lifecycle typically comprises the following phases³⁵⁷: (i) planning, (ii) data understanding, (iii) data preparation, (iv) modelling, (v) evaluation, and (vi) deployment and monitoring. These stages evidently require continuous engagement between the provider of the AI system, who bears responsibility for its design and development, and the deployer, who is responsible for implementing the system within a specific context and for a defined purpose. The deployer, therefore, ought to be involved throughout the entire process, as it is best positioned to define the analytical objectives and to exercise the value judgments that such objectives inevitably entail.

The intersection between the respective and equally crucial roles of the provider and the deployer within the AI lifecycle – representing, in my analysis a key element in addressing issues such as bias and discrimination – is, however, difficult to achieve in practice. One primary reason for this challenge lies in the fact that AI systems used by public authorities are frequently procured from private third-party developers³⁵⁸. In the worst-case scenario – though regrettably the most frequent one – such developers are large corporations possessing significant bargaining power. Within the specific context of law enforcement, for instance, Palantir has supplied its products to numerous European law enforcement authorities. Consequently, public bodies typically have limited or no control over the data used for training, validation, and testing. Yet, since the data employed must be representative of the population to which the AI system will be applied, a further concern arises. Public administrations often hold extensive datasets – frequently encompassing sensitive or personal

³⁵⁶ This addition has been considered necessary in the literature as otherwise the obligations would have arguably been an obligation too difficult to comply with. On this point see See M. EBERS, V. R. HOCH, F. ROSENKRANZ, H. RUSCHEMEIER, B. STEINRÖTTER, *The European Commission's proposal for an Artificial Intelligence Act - A critical assessment by members of the robotics and AI law society (RAILS)*, in *J*, 2021, 4, 589–603.

³⁵⁷ C. SHEARER, *The CRISP-DM model: the new blueprint for data mining*, in *Journal of Data Warehousing*, 5(4), 2000, 13–22. VARSHNEY, *Trustworthy Machine Learning*, 14.

³⁵⁸ MARCHETTI, *Amministrazione digitale*, cited, 100.

information, particularly in the case of police data – which may need to be shared with private providers for the development or training of AI models. Such data transfers raise significant privacy and data protection concerns, particularly regarding compliance with data protection legislation and the safeguarding of citizens’ rights. These issues can, to a considerable extent, be mitigated when AI systems are developed in-house by public authorities. In this scenario, the administration retains full control over the datasets used for training, validation, and testing, ensuring that data remain within the public domain and is not transmitted to external entities. Pursuant to Article 26(4) of the AI Act, where deployers exercise control over the data fed into a high-risk AI system, they are required to ensure that such data is “*relevant, accurate, and sufficiently representative*” for the intended purpose of the AI system. Therefore, where the development of AI systems in-house is feasible, this option ought to be preferred, as it allows deployers – particularly when they are public authorities in sensitive domains – to retain greater control over design choices, and data and data governance³⁵⁹.

The mere fact that the in-house development of an AI system is not feasible due to a lack of technical capabilities or resources does not absolve public authorities of their responsibilities. When public authorities are not themselves the developers of an AI system, they nonetheless remain accountable for ensuring effective oversight of the data employed in its development. Failing to do so may lead to significant accountability gaps, as illustrated by the English case *R (Bridges) v. Chief Constable of South Wales Police*³⁶⁰. In that instance, the Court found the police force in breach of the Public Sector Equality Duty (PSED) precisely because it lacked access to the datasets used by the provider of the face recognition technology. Ensuring such control requires that appropriate contractual clauses be negotiated and included in procurement agreements to guarantee access to, and control over, the datasets used for training, validation, and testing.

The obligations enshrined in Article 10 are highly complex yet fundamental to the proper functioning of AI systems, as they serve to prevent – or, at the very least, to mitigate – the risk of algorithmic outputs that are flawed, biased, or discriminatory. In this instance as well, the by design approach is clearly reflected, as AI systems must be conceived and developed in a manner that seeks to prevent or, where avoidance is not possible, to mitigate potential discriminatory outcomes. In particular, the representativeness of data appears to be a crucial factor in mitigating bias³⁶¹. Article 10(3) of the AI Act explicitly requires that training, validation, and testing datasets be, among other things,

³⁵⁹ MARCHETTI, *Amministrazione digitale*, cited, 100. MARCHETTI, TORCHIA, *AI and public administration*, cited.

³⁶⁰ UK Court of Appeal (Civil division), Judgement of 11 August 2020, *Ed Bridges v. The Chief Constable of South Wales Police and others*. The Court of Appeal, overturning the judgment of the High Court, found the use of automated facial recognition (AFR) technology to be unlawful on three grounds: (i) Article 8 of the ECHR, (ii) the Data Protection Act 2018, and (iii) the Public Sector Equality Duty under section 149 of the Equality Act 2010.

³⁶¹ In similar terms see R. XENIDIS, *Tuning EU equality law to algorithmic discrimination: Three pathways to resilience*, in *Maastricht Journal of European and Comparative Law*, 27(6), 2020, 736-758.

“*sufficiently representative*”. The representativeness of such data must be evaluated with reference to the specific context in which the AI system is intended to be deployed, ensuring that the data reflect the relevant characteristics of the population and environment to which the system will apply. Specifically, datasets must, insofar as required by the AI system’s intended purpose, reflect the characteristics or elements particular to the relevant geographical, contextual, behavioural, or functional environment within which the high-risk AI system will operate³⁶². Although this obligation is formally directed at providers of high-risk AI systems, a corresponding duty is likewise imposed on deployers (Article 26)³⁶³.

As discussed in Chapter 2, police data may suffer from unrepresentativeness, meaning that the information collected and recorded by law enforcement authorities does not necessarily reflect the true extent or nature of criminal activity within a population. Several factors can contribute to this issue. For example, police data typically include only incidents that are reported to or detected by the police, thereby excluding offences that go unreported or unrecorded. Moreover, patterns of policing, such as where officers are deployed, the communities they engage with most frequently, and the types of behaviours they prioritize, can also distort the data they collect. As a result, certain groups or neighbourhoods may appear to have higher crime rates simply because they are subject to more intensive surveillance or enforcement activity, not necessarily because more crime actually occurs there. Therefore, while police data are valuable for understanding trends and operational needs, they must be interpreted with caution and supplemented by other sources, such as victimisation surveys or self-report studies, to build a more accurate and comprehensive picture of crime. Where law enforcement bodies fail to scrutinise and validate their datasets, or those supplied by the provider, the risk of reinforcing systemic biases is significantly heightened.

Ensuring fairness in data used for AI training and deployment represents a convergent obligation arising from both AI-specific regulation and data protection law whenever personal data are processed. Although data protection legislation does not furnish a precise technical definition of fairness, the principle nonetheless constitutes a fundamental component of EU data protection law, grounded in Article 8(2) of the CFREU and operationalised through the GDPR and LED’s requirement for fair processing. In particular, the European Data Protection Board has interpreted fairness as an overarching principle mandating that personal data are not processed in ways that are unjustifiably detrimental, unlawfully discriminatory, unexpected, or misleading to data subjects³⁶⁴. This multidimensional conception addresses several distinct fairness concerns, namely protection

³⁶² AI Act, Article 10(4).

³⁶³ AI Act, Article 26(4).

³⁶⁴ European Data Protection Board, *Guidelines on Data Protection by Design and by Default*, 20 October 2020, paragraph 69. Available at: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en.

against unjustified harm; prohibition of discriminatory treatment (whether direct or indirect through disparate impact); alignment with data subjects' reasonable expectations based on collection context; and transparency regarding actual data uses. These dimensions collectively establish that fairness extends beyond mere legal compliance to encompass substantive equity.

Therefore, for law enforcement authorities functioning simultaneously as data controllers and AI providers or deployers, these fairness obligations translate into concrete requirements to identify, measure, and mitigate biases throughout the AI lifecycle. Identification involves examining training data for representational imbalances, proxy discrimination, and encoded historical biases, recognising that law enforcement data frequently reflects past discriminatory practices that, if uncorrected, become automated through AI systems. Measurement requires selecting and applying appropriate fairness metrics, navigating the inherent tensions between different fairness definitions (demographic parity, equalized odds, individual fairness) while recognising that metric choice depends on context, affected rights, and potential harms. Mitigation encompasses technical interventions (data preprocessing, algorithmic constraints, output adjustments) as well as procedural and institutional safeguards ensuring ongoing fairness monitoring. These obligations carry particular weight in law enforcement given the domain's characteristics I discussed throughout the present Thesis. Law enforcement authorities must therefore approach fairness not as a compliance checkbox but as an ongoing governance commitment requiring sustained technical, procedural, and institutional efforts to prevent AI systems from perpetuating or amplifying societal inequalities.

Furthermore, recognising that bias detection and correction constitute a substantial public interest directed at protecting natural persons from adverse effects, including discrimination, the European legislator has taken an additional step by expressly permitting, under certain conditions, the processing of special categories of personal data by providers of high-risk AI systems³⁶⁵. Such processing is legally authorised only insofar as it is strictly necessary for ensuring the effective detection and correction of bias within those systems. Accordingly, Article 10(5) establishes the legitimate basis under Union law for processing special categories of personal data, as required by Article 9 of the GDPR and Article 10 of the LED³⁶⁶. Moreover, acknowledging that discrimination

³⁶⁵ AI Act, Article 10(5).

³⁶⁶ In addition to the conditions set out in Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680, all the following conditions must be met in order for such processing to occur: (a) the bias detection and correction cannot be effectively fulfilled by processing other data, including synthetic or anonymised data; (b) the special categories of personal data are subject to technical limitations on the re-use of the personal data, and state-of-the-art security and privacy-preserving measures, including pseudonymisation; (c) the special categories of personal data are subject to measures to ensure that the personal data processed are secured, protected, subject to suitable safeguards, including strict controls and documentation of the access, to avoid misuse and ensure that only authorised persons have access to those personal data with appropriate confidentiality obligations; (d) the special categories of personal data are not to be transmitted, transferred or otherwise accessed by other parties; (e) the special categories of personal data are deleted once the bias has been corrected or the personal data has reached the end of its retention period, whichever comes first; (f) the records of processing activities pursuant to Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680

may arise from AI systems other than high-risk ones, the European Commission has recently proposed extending this legal basis to providers and deployers of other AI systems and AI models, as well as to deployers of high-risk AI systems³⁶⁷.

Having data of high quality is vital also for the technical reliability of the AI system, thus provisions concerning data are crucial for ensuring the production of accurate outcomes³⁶⁸. To this end, the performance of an AI model must be carefully evaluated³⁶⁹. There are various metrics available to assess whether a model performs effectively, among which the most common are accuracy, precision, and recall³⁷⁰. The choice of metric depends on the specific context in which the model operates. For instance, precision is an appropriate measure when the cost of false positives is high – such as when a false positive could lead to the wrongful incarceration of an individual – whereas *recall* is more relevant when false negatives carry greater risks, for example, when failing to identify a potential terrorist³⁷¹. More generally, both the High Level Expert Group³⁷² and the European legislator³⁷³ have adopted the concept of “accuracy” in a broad and more flexible sense. The accuracy of an AI system is fundamentally determined by its development process and, more specifically, by the nature and quality of the data employed during training³⁷⁴. The quality of the data used to train an algorithm – particularly its representativeness – is crucial to ensuring the accuracy of AI models. For example, in the case of RBI systems, deficiencies in training data may impair the system’s ability to correctly identify individuals from certain demographic groups, resulting in disproportionate and unfair impacts on those individuals. Similarly, low-quality data – such as data collected under uncontrolled conditions – can adversely affect the accuracy of AI models, leading to a higher incidence of false results.

In conclusion, although the provisions concerning data are broadly positive, their implementation requires careful attention and involves considerable complexity for both providers and deployers. The Parliament, in its amendments, introduced the possibility of “outsourcing” the requirements related

include the reasons why the processing of special categories of personal data was strictly necessary to detect and correct biases, and why that objective could not be achieved by processing other data.

³⁶⁷ European Commission, Proposal for a Regulation Of The European Parliament And Of The Council amending Regulations (EU) 2024/1689 and (EU) 2018/1139 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI), Brussels, 19.11.2025, COM(2025) 836 final, Recital 6.

³⁶⁸ AI Act, Article 15.

³⁶⁹ For an interesting discussion on how to evaluate the performance of AI models see UNICRI, INTERPOL, *Responsible AI Innovation in law enforcement. AI toolkit - 07. Technical Reference Book*, cited, 30. The performance “is based on how close the predictions were to the actual, measured values”.

³⁷⁰ *Ibidem*.

³⁷¹ *Ibidem*.

³⁷² HLEG, *Ethics guidelines for trustworthy AI*, cited, 17.

³⁷³ AI Act, Article 15.

³⁷⁴ UNICRI, INTERPOL, *AI Toolkit – 0.2. Principles for Responsible AI Innovation*, cited, 12. B. HANSON ET AL., *Garbage in, garbage out*, cited.

to data governance³⁷⁵, acknowledging that the extensive workload imposed by Article 10 could place a significant burden on providers and potentially hinder innovation. Building on this development, some authors have elaborated the notion of new actors responsible for verifying and certifying that data comply with the established governance requirements, the so-called “data verifiers” or “certified data service providers”³⁷⁶. In any event, the inclusion of the phrase “to the best extent possible” has arguably rendered compliance with the obligation under Article 10 more technically achievable. Nevertheless, as with the transparency requirements, the AI Act refrains from establishing concrete criteria for evaluating dataset quality, instead delegating such specifications to the domain of standardisation³⁷⁷. Ultimately, the implementation of a standardized process for managing data quality may prove particularly advantageous within law enforcement, as it facilitates the development of a systematic, step-by-step approach applicable to all operational activities.

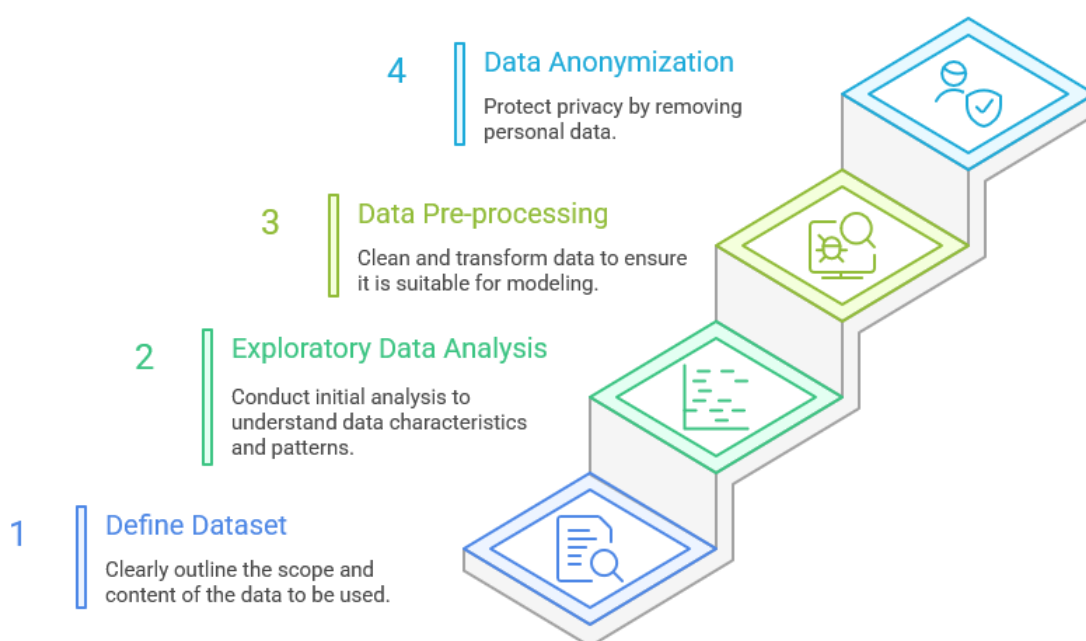


Figure 15: Steps to prepare data for modeling based on UNICRI, INTERPOL, *Responsible AI Innovation in law enforcement. AI toolkit - 07. Technical Reference Book*. Own elaboration with Napkin.

³⁷⁵ Amendment 78 modifying Recital 44.

³⁷⁶ CORERA, *Data And Data Governance And Connections To Data Protection Principles In Article 10 Of The Artificial Intelligence Act*, cited, 611. At the same time the author critically look at such introduction by stating that “is questionable whether this model of “verified” data provision can deliver the individualised compliance with such governance requirements that the Regulation aspires to, since, firstly, governance must be tailored to the context of use as well as to the intended purpose of the AI system and, secondly, the sets of data sets that will be used for the purpose of the AI system will need to be defined, secondly, datasets should take into account, to the extent required by the intended purpose, the characteristics or elements specific to the geographical, behavioural, contextual, or functional environment in which the high-risk AI system is intended to be used”.

³⁷⁷ CORERA, *Data And Data Governance And Connections To Data Protection Principles In Article 10 Of The Artificial Intelligence Act*, cited, 615.

3.3. *The role of human oversight in mitigating algorithmic discrimination*

The AI Act positions human oversight as one of its central regulatory mechanisms for the prevention and mitigation of algorithmic discrimination, embedding it within a multi-layered governance architecture that spans the entire lifecycle of high-risk AI systems. As discussed in previous Sections, fairness concerns are broad, multifaceted, and deeply complex, with technical, legal, ethical, and organisational dimensions. Achieving fairness in practice therefore necessitates sustained normative and contextual scrutiny across the entirety of the decision-making process, from data collection and model training, through threshold-setting and classification, to the human determinations that follow from and act upon AI systems' outputs³⁷⁸, rather than at any single, isolated stage.

It is at the final, consequential stage, where model predictions are translated into tangible interventions, however, that discriminatory dynamics may most acutely materialise. A risk score, a classification, or a recommendation, however technically refined, carries no inherent normative authority; its significance is constituted by the institutional and human context in which it is received and acted upon. The requirement of meaningful human oversight accordingly assumes a central and indispensable role in mitigating the risk that algorithmic outputs serve as the uncritical basis for decisions with potentially discriminatory effects, particularly in domains, such as law enforcement, where fundamental rights are directly at stake.

As established in the preceding analysis, human oversight under the AI Act is directed at preventing or minimising risks to fundamental rights, including the risk of discrimination, and requires that natural persons assigned oversight responsibilities be enabled to perform a series of critical functions. Specifically, Article 14 mandates that human overseers be equipped to: (a) interpret the high-risk AI system's output in a manner that is both technically informed and contextually sensitive; (b) exercise a genuine decisional prerogative, including the capacity to decline to use the system, to disregard its output, or to override or reverse its recommendations; and (c) intervene in the operation of the high-risk AI system when circumstances so require.

These requirements are further elaborated by Recital 73, which specifies that high-risk AI systems shall incorporate "*in-built operational constraints that cannot be overridden by the system itself*", that they be "*responsive to the human operator*", and that oversight personnel possess "*the necessary competence, training and authority*" to discharge their oversight function effectively. The combined effect of Article 14 and Recital 73 is to establish a normative expectation that human oversight constitutes not a mere formal checkpoint, but a substantive mechanism of output control, error correction, and bias mitigation, one that is structurally integrated into the operational deployment of

³⁷⁸ LAGIOIA ET AL., *Algorithmic fairness through group parities?*, cited.

the AI system and sustained throughout its lifecycle. Article 26 concludes the cycle by requiring deployers of high-risk AI systems to exercise effective human oversight over those systems throughout their operation³⁷⁹. In doing so, it extends the oversight obligation beyond the design and development phases – where providers bear primary responsibility – to the point of deployment, ensuring that the actors most proximate to the system’s real-world impact bear a continuing duty to monitor, evaluate, and, where necessary, intervene in its functioning.

From a careful reading of these provisions, human oversight should be understood as performing several interconnected functions within the broader regulatory architecture³⁸⁰. First, it serves as a direct safeguard for output control and the mitigation of risks to fundamental rights, including systematic biases, in AI-supported decisions. Second, the requirement of human oversight is intimately linked to questions of accountability and responsibility. By designating identifiable natural persons as overseers, the AI Act facilitates the allocation of responsibility for the consequences of AI-assisted decisions. This ensures that actors deploying high-risk AI systems – particularly professional actors such as law enforcement authorities – can be held answerable for the impact of those systems and the fairness of the outcomes they produce. This is of particular importance in the public sector, where fairness constitutes a foundational principle of public law, and where public deployers bear a heightened obligation to ensure that the decisions they make comply with the legal principle of non-discrimination. Third, human oversight possesses a distinctive capacity to introduce contextual sensitivity into the decision-making process. Algorithmic systems, by their nature, operate on the basis of generalised patterns extracted from the data on which they are trained. They are structurally incapable of apprehending the full complexity of individual circumstances, the particularities of intersecting vulnerabilities, or the normative significance of contextual factors that may render a formally equal classification substantively unjust. Human overseers, by contrast, are – at least in principle – positioned to exercise situated judgment: evaluating complex real-world scenarios and identifying the most appropriate course of action. In principle, they can recognise when an algorithmically generated prediction may produce disproportionate or unjust effects in a specific case and modulate or override the system’s output accordingly. Fourth, the strength of the AI Act’s

³⁷⁹ The AI Act thus introduces “a shared responsibility between the developer, who provides for their AI to be overseable, and the user/deployer, who executes the oversight”. J. LAUX, *Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act*, in *AI & Society*, 39, 2024, 2855.

³⁸⁰ G. LAZCOZ MORATINOS, *Human Oversight Or Monitoring In Article 14 Of The Artificial Intelligence Act: A Mere Mandatory Requirement For High-Risk Systems?*, in L. COTINO HUESO, D.-U. GALETTA (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, cited, 717-737. The relevance of human oversight has been specifically highlighted in relation to high-stake decisions, such as in medical decision-making and in judicial contexts. See I. BANKS, *Judges-in-the-loop? Judicial involvement in human oversight of high-risk decision support systems under the EU AI Act*, in *International Journal of Law and Information Technology*, 34, 2026. S. KALTENBRUNNER, *Human in control: Shared decision-making with clinical decision-support systems under the Artificial Intelligence Act*, in *Computer Law & Security Review*, 61, 2026, 106281.

human oversight requirement lies in its lifecycle approach to oversight. Rather than confining the oversight obligation to a single temporal moment, such as the point of deployment or the moment of individual decision-making, the provision contemplates oversight as a continuous and adaptive process, encompassing the monitoring of system performance, the identification of emerging risks, and the ongoing evaluation of the system's interaction with the populations it affects. This lifecycle orientation is essential for the detection of discriminatory dynamics that may not be apparent at the outset but emerge over time as the system encounters new data distributions, shifting social contexts, or previously unrecognised patterns of disparate impact.

Notwithstanding the normative ambition of the AI Act provisions on human oversight, a substantial and growing body of scholarship has subjected the effectiveness of oversight to sustained critical scrutiny. The central concern articulated in this literature is that the AI Act systematically overestimate the capacity of human actors to exercise meaningful, independent, and critical oversight over AI systems, especially when they are particularly complex. Human cognition is subject to well-documented constraints that are particularly relevant in the context of AI oversight. Bounded rationality³⁸¹, information overload, time pressure, decision fatigue, and the inherent difficulty of evaluating probabilistic and high-dimensional outputs all operate to diminish the quality and reliability of human evaluative judgments when confronted with AI systems' predictions. These limitations are not incidental or remediable through modest procedural adjustments; they are constitutive features of human cognitive architecture.

The most significant empirical challenge to the efficacy of human oversight in mitigating algorithmic discrimination is the phenomenon of automation bias, discussed in Section 3.3., representing the tendency of human decision-makers to defer excessively to automated predictions, even in circumstances where independent judgment would lead to a different conclusion. Automation bias inevitably reduces the likelihood that human overseers will identify and correct errors in algorithmic output – including discriminatory ones – thereby converting what is intended as a safeguard into what may function, in practice, as a mechanism of mere ratification. As discussed earlier, the AI Act introduces an awareness of this tendency.

By contrast, it does not require an awareness of the opposing tendency to override AI outputs even when they are highly reliable – so-called algorithmic aversion³⁸², due to distrust of the system or an inability to interpret it – or of the risk of exercising one's discretion in systematically biased ways –

³⁸¹ See *Bounded Rationality*, in *Stanford Encyclopedia of Philosophy*, First published 30 November 2018; substantive revision 13 December 2024. Available at: <https://plato.stanford.edu/entries/bounded-rationality/>.

³⁸² B. BOTERO ARCILA, *AI liability in Europe: How does it complement risk regulation and deal with the problem of human oversight?*, in *Computer Law & Security Review: The International Journal of Technology Law and Practice*, 54, 2024, Article 106012, 4.

so-called selective adherence³⁸³. Selective adherence poses an especially acute problem for the detection of discriminatory AI-based decisions. As Alon-Barkat and Busuioc found, overseers are disproportionately inclined to defer to algorithmic recommendations precisely when those recommendations reinforce pre-existing stereotypes, implicit biases, or prior assumptions about particular social groups. Under these conditions, the algorithmic output ceases to function as an independent assessment subject to critical scrutiny and instead operates as an authoritative seal of approval on the overseer's own prejudices, compounding the very discriminatory dynamics that oversight was designed to counteract.

The findings discussed above carry significant implications for the regulatory design of human oversight under the AI Act. They suggest that, under certain conditions, interposing a human overseer may prove not merely ineffective but paradoxically counterproductive, amplifying rather than attenuating the discriminatory effects of an algorithmic system. These concerns, especially acute in high-stake contexts, find empirical support in a large-scale experiment conducted by the European Commission's Joint Research Centre, involving 1,411 human resources and banking professionals across Italy and Germany³⁸⁴. The research shows that human overseers exhibited no statistically significant difference in their propensity to follow the advice of a fair AI system as compared with a demonstrably discriminatory AI system. This result suggests that the capacity of human overseers to detect and resist discriminatory algorithmic outputs may be far more limited than the regulatory architecture of the AI Act presupposes, raising fundamental questions about the conditions under which human oversight can be expected to function as an effective anti-discrimination mechanism. The findings above thus lead to a sobering conclusion: human oversight, while necessary, is demonstrably insufficient as a standalone safeguard against algorithmic discrimination. The AI Act itself appears to acknowledge it, constructing an integrated regulatory architecture in which human oversight interacts with and is reinforced by several other substantive requirements, each of which independently and collectively targets the risk of algorithmic discrimination.

The first of these complementary provisions is Article 9, which establishes a continuous, lifecycle-spanning risk management system, as previously discussed. This provision requires providers and, where applicable, deployers of high-risk AI systems to identify and analyse “*known and reasonably*

³⁸³ S. ALON-BARKAT, M. BUSUIOC, *Human–AI Interactions in Public Sector Decision Making: “Automation Bias” and “Selective Adherence” to Algorithmic Advice*, in *Journal of Public Administration Research and Theory*, 33, 2023, 153.

³⁸⁴ A. GAUDEUL, O. ARRIGONI, V. CHARISI, M. ESCOBAR-PLANAS, I. HUPONT, *Understanding the Impact of Human Oversight on Discriminatory Outcomes in AI-Supported Decision-Making*, in *ECAI*, 392, 2024, 1067-1074. Joint Research Centre, A. GAUDEUL, O. ARRIGONI, V. CHARISI, M. ESCOBAR-PLANAS, I. HUPONT, *The Impact of Human-AI Interaction on Discrimination. A large case study on human oversight of AI-based decision support systems in lending and hiring scenarios*, Publications Office of the European Union, Luxembourg, 2025. Available at : <https://publications.jrc.ec.europa.eu/repository/handle/JRC139127>.

foreseeable risks” to fundamental rights, including, explicitly, discrimination risks, with specific and heightened attention to potential impacts on vulnerable groups as specified in Article 9(9). The discrimination risks identified through this systematic process directly inform the nature, intensity, and proportionality of the human oversight measures required under Article 14(3), ensuring that oversight arrangements are commensurate with the specific discrimination risks posed by a given system in each deployment context. Second, Article 10 on data governance represents what is arguably the most explicit anti-discrimination provision in the AI Act. By imposing stringent requirements on the quality, representativeness, and appropriateness of training, validation, and testing datasets, it addresses one of the most significant upstream sources of algorithmic discrimination: the encoding of historical patterns of inequality and exclusion in the data upon which AI models are trained. To the extent that robust data governance succeeds in reducing the discriminatory content of training data, the burden placed on downstream human oversight to detect and correct biased outputs is correspondingly diminished, though, as the empirical evidence reviewed above suggests, by no means eliminated. Third, Article 13 on transparency establishes a prerequisite condition for the very possibility of effective human oversight, as discussed in Part I. By requiring that high-risk AI systems be “*sufficiently transparent to enable deployers to interpret a system’s output and use it appropriately*”, Article 13 seeks to ensure that the informational preconditions for meaningful oversight are satisfied. Without adequate transparency regarding the factors, variables, and reasoning processes that contribute to an algorithmic output, human overseers are structurally incapable of detecting discriminatory patterns, evaluating the reliability of system predictions, or exercising the informed evaluative judgment that Article 14 contemplates. Fourth, Article 27 further targets the risk of discrimination by mandating that deployers of high-risk AI systems conduct a FRIA prior to the deployment of such systems. As discussed in Section 3.1., the FRIA obligation compels deployers to undertake a structured, *ex ante* evaluation of the potential discriminatory effects of the AI system in its intended operational context, thereby ensuring that fairness considerations are integrated into the decision to deploy a system before any individual is affected by its outputs.

The path forward thus requires a decisive shift in the conceptualisation of human oversight, moving beyond a model centred on individual cognitive performance toward one grounded in institutionalised, systemic governance. This reconceptualization entails the development of integrated oversight ecosystems that combine meaningful human review with algorithmic auditing, specialised anti-discrimination training for overseers and democratic accountability mechanisms. The findings of the European Commission’s Joint Research Centre study point to the need for precisely such a shift, from an oversight model that places the cognitive and moral burden of bias detection on individual human actors to an integrated system designed to address both human cognitive limitations

and the discriminatory tendencies of AI systems in a mutually reinforcing manner. This systemic approach recognises that the problem of algorithmic discrimination is not primarily a problem of individual failure but of institutional design: it arises from the interaction between technical systems, organisational practices, informational environments, and human cognitive architecture, and it must be addressed at the level of the system as a whole rather than through reliance on the vigilance of individual overseers.

To translate this systemic vision into operational reality, several practical instruments may be taken into consideration. First, overseers require access to interpretability tools and decision-support frameworks that render AI outputs legible, through adequate tools that highlight the factors driving a particular recommendation³⁸⁵, flag cases in which the system's confidence is low or its output deviates from established patterns, and provide contextual information relevant to the assessment of fairness in individual cases. Second, organisational protocols may be established that structure the oversight process in ways that actively counteract the documented effects of automation bias and selective adherence. Such protocols might include mandatory independent assessment requirements, for instance obligating overseers to formulate a preliminary judgment before consulting the algorithmic output, structured decision-recording procedures that require overseers to articulate the reasoning behind their determinations, and periodic auditing of oversight decisions to identify patterns of uncritical deference or discriminatory selectivity. Third, training programmes for human overseers must extend beyond technical literacy to encompass substantive engagement with the dynamics of algorithmic discrimination, the psychology of automation bias and selective adherence, and the specific legal obligations arising under anti-discrimination law. Moreover, the elaboration of harmonised standards, despite the limitations discussed above, may present a decisive opportunity to operationalise the requirements of Article 14 in a manner that is informed by the empirical evidence on automation bias, selective adherence, and the conditions under which human oversight can meaningfully contribute to the mitigation of algorithmic discrimination. Enforcement practices, too, will play a constitutive role in determining the practical significance of the AI Act's oversight requirements. The extent to which market surveillance authorities scrutinise the adequacy of human oversight arrangements, assessing not merely whether oversight is formally in place, but whether it is substantively effective in detecting and preventing discriminatory outcomes, will shape the incentive structures facing providers and deployers, and ultimately determine whether the AI Act's requirement of human oversight is realised in operational practice or remains a largely aspirational norm.

³⁸⁵ See Part I, Sections 2.3 and 2.4.

CONCLUDING REMARKS

Police authorities collect vast amounts of data on a daily basis. These data are stored in large databases and require continuous analysis to inform operational and strategic decisions. It is widely acknowledged that the use of digital technologies – of which AI represents the current frontier – can support law enforcement authorities in extracting meaningful insights and achieving a more comprehensive understanding of the information they gather, thereby enhancing the effectiveness and efficiency of their functions.

Throughout this Thesis, I examined the concept of AI and its relevance to law enforcement, explored some of the most contentious and high-risk applications of AI in this context, and analysed the specific provisions that the AI Act introduces for AI systems deployed for law enforcement purposes. I then focused on two essential aspects for ensuring public trust and acceptance in this sensitive domain: transparency and fairness, which both are necessary for law enforcement authorities to carry out their functions effectively. The central research question guiding this Thesis is whether public values – namely, the protection of fundamental rights, transparency, fairness, the rule of law – are adequately protected when law enforcement authorities carry out their functions through the use of AI systems. The concluding section addresses this question directly, drawing together the findings of the analysis and highlighting the key results of the research.

Although the overall structure of the AI Act is coherent and well-conceived, and the inclusion of explicitly prohibited practices together with detailed requirements and obligations for high-risk AI systems represents a significant advancement, certain shortcomings in the protection of public values and fundamental rights nonetheless persist. This is, however, not entirely unexpected, as the AI Act establishes a minimum harmonisation framework that aims to set a sufficiently high, yet not exhaustive, standard of protection across the European Union. In the law enforcement sector, a pertinent example is Italian Law No. 132/2025, which under Article 24(2)(h) delegates to the Government the power to adopt a legislative decree specifically governing the use of AI systems in that context.

Three overarching considerations guide the analysis in this final section. First, the AI Act does not operate in isolation. It must be understood as part of a broader legal ecosystem, most notably alongside European data protection legislation. The GDPR and, specifically, the LED continue to govern how personal data may be collected and processed within AI systems, and they significantly influence how AI systems can be deployed in practice. Second, human rights law, and the

CONCLUDING REMARKS

corresponding jurisprudence, provide essential guidance for interpreting the limits of permissible state conduct. They help clarify where the boundaries lie in terms of surveillance, and interference with fundamental rights, as well as what procedural and substantive safeguards are required to protect individuals from disproportionate or unjustified state action. Third, the AI Act introduces a highly intricate set of classifications, obligations, and compliance mechanisms. Due to its complexity, effective implementation requires substantial organisational adaptation. To ensure that AI systems are used lawfully, transparently, and in accordance with the AI Act's requirements, law enforcement authorities will need new governance structures or to revise existing ones, as well as the necessary expertise and operational routines.

1. A final assessment: is the AI Act sufficiently adequate in protecting rights and public values?

Throughout the Chapters, I analysed the AI Act, its structure, and the safeguards it introduces for individuals when law enforcement activities are augmented by AI. Despite an overall positive assessment, several questions remain. Is the definition of an AI system sufficiently clear and comprehensive? Is the definition of law enforcement precise enough to determine with certainty when a given activity falls within its scope? And does the risk-classification framework adequately account for the variety of AI systems used for law enforcement purposes?

As for the first question, the AI Act applies only to “AI systems”¹. As discussed in Chapter 1, a system qualifies as an AI system under this provision if it meets the characteristics set out in Article 3 of the AI Act. Despite the Commission's *Guidelines on the definition of an artificial intelligence system*, significant uncertainty still surrounds the definition. This becomes clear when examining the cases discussed in Chapter 2, which illustrate how difficult it can be to determine whether a given system does – or does not – fall within the scope of the definition. As shown in that Chapter, not all predictive policing systems are alike, and their technical characteristics vary considerably. Determining whether a given system falls within the definition enshrined in Article 3 remains open to debate. The task is even more complex for software used by law enforcement authorities, where publicly available information is scarce, making it difficult to analyse the system's specific characteristics and reach a well-grounded assessment.

¹ See also European Commission, *ANNEX to the Communication to the Commission. Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)*, Brussels, 6.2.2025 C(2025) 924 final ANNEX. See also European Law Institute, *The concept of 'AI system' under the new AI Act: Arguing for a Three-Factor Approach*, December 2024.

CONCLUDING REMARKS

Moreover, the current definition of AI systems still excludes certain types of software used for predictive policing, which remain classified as traditional software. Although the systems falling outside the definition may not be particularly complex or present the greatest risks to individuals, they can nevertheless produce biased outcomes, contributing to the over-policing of particular areas and specific communities, with inevitable consequences in terms of feedback loops and potential discrimination. Similar considerations apply to automated video surveillance discussed in Chapter 2. Most contemporary video-analytics systems described as “intelligent” fall within the AI Act’s definition of AI system. This is the case whenever the technology infers information from video data, rather than merely applying simple, fixed rules. For example, automated surveillance may be considered an AI system when cameras are equipped with tools capable of recognising suspicious behaviour, detecting crowd movements or loitering, identifying fights or abandoned objects, signalling intrusions into restricted areas, performing face recognition, categorising individuals based on facial or bodily features, tracking and re-identifying persons, or detecting recurrent behavioural patterns and generating corresponding alerts. Only very simple rule-based systems (such as basic motion detection or fixed sensor-triggered alarms) appear to fall outside this definition, remaining classified as traditional software.

Once it is established that a given solution falls within the definition of an AI system, the subsequent classification of its risk level becomes crucial. Assessing whether the AI Act adequately protects individuals and public values in the context of AI-enhanced policing requires a nuanced evaluation. On the one hand, the AI Act represents a significant step forward: it introduces prohibitions targeting the most intrusive practices, such as some forms of predictive policing and biometric recognition, and establishes a comprehensive regulatory regime for high-risk systems, embedding requirements and obligations relating to risk management, data, transparency, fairness. These mechanisms collectively aim to mitigate risks of discrimination, unwarranted surveillance, and transparency concerns that are particularly acute in the law enforcement domain.

On the other hand, important limitations remain. The degree of protection afforded to individuals varies according to the risk category into which the AI system is classified. When the AI system falls within the category of prohibited AI practices, individuals benefit from the strongest level of protection, as such systems are, in principle, barred from development and/or deployment. As discussed in Chapter 3, however, the scope of prohibited practices is relatively narrow, with multiple exceptions that raise the risk of “*function creep*”² whereby technologies initially introduced under narrow and specific conditions gradually expand into broader applications, undermining the intended

² N. A. SMUHA, K. YEUNG, *The European Union’s AI Act: beyond motherhood and apple pie?*, in N.A. SMUHA (Ed.), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*, Cambridge Law Handbooks, Cambridge University Press, 2025, 237.

CONCLUDING REMARKS

safeguards for fundamental rights and freedoms. Regarding high-risk AI systems, the AI Act introduces specific requirements and obligations primarily designed to mitigate and manage risks, that can emerge in different phases of the AI lifecycle, to an acceptable level. Beyond the conceptual ambiguity surrounding what constitutes an “acceptable” level of risk, these requirements prove challenging to operationalise in practice, necessitating both normative interpretation of the AI Act’s provisions and substantial technical expertise. To address this implementation gap, the AI Act mainly relies upon technical standards and common specifications as supporting instruments. While this reliance on standards presents certain drawbacks discussed in Chapter 3, the most significant challenge is the current absence of harmonised standards that provide concrete guidance on compliance with the AI Act’s high-risk requirements. In response to this deficiency, the European Commission has recently proposed linking the implementation timeline of high-risk rules to the availability of standards and support tools, rather than to fixed deadlines³.

Particular attention should be paid to Article 6(3) of the AI Act, which provides that AI systems that would generally qualify as high-risk under the use cases listed in Annex III may nonetheless be exempt from high-risk requirements where they do not pose a “*significant risk*” to health, safety, or fundamental rights. Although AI providers are required to register such decisions in an EU database pursuant to Article 6(4) of the AI Act, Article 49(4) introduces exceptions in the area of law enforcement that limit the transparency of these registrations. In any case, the practical applicability of the filter will ultimately depend on how the Commission elaborates upon these criteria and how the CJEU interprets them.

Beyond these implementation challenges, determining the appropriate risk category itself proves problematic. For instance, the distinction between predictive policing systems that qualify as prohibited practices and those classified as high-risk systems hinges on whether the system relies *solely* on the profiling of natural persons. Systems that use profiling as their exclusive basis fall under the prohibition, while those that incorporate additional methodologies may instead be deemed high-risk. Therefore, this distinction is complicated by two factors: first, the inherent ambiguity in defining what constitutes “profiling”⁴; and second, the critical importance of interpreting the adverb “solely”, which becomes decisive given that different classifications carry significant legal and practical consequences. An illustrative example of a predictive policing system based solely on profiling may

³ European Commission, Proposal for a Regulation Of The European Parliament And Of The Council amending Regulations (EU) 2024/1689 and (EU) 2018/1139 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI), Brussels, 19.11.2025, COM(2025) 836 final.

⁴ For detailed discussion on the topic see EDPB, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, 2018. Information Commissioner’s Office, *What is automated individual decision-making and profiling?*. Available at: <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/individual-rights/automated-decision-making-and-profiling/what-is-automated-individual-decision-making-and-profiling/>.

CONCLUDING REMARKS

be an AI tool deployed by law enforcement authorities to forecast the likelihood that specific individuals will engage in future criminal activity; the system does not integrate contextual or situational data – such as local crime trends, environmental conditions, or spatial-temporal patterns of offending – instead, it generates individual risk assessments exclusively on the basis of personal attributes and behavioural indicators associated with those individuals, for example, the algorithm may rely on variables including age, gender and ethnicity; these inputs form comprehensive personal profiles from which the system infers a probability of future offending and the resulting risk scores are then used to inform policing decisions, such as the allocation of surveillance resources or the initiation of preventative interventions, without human involvement. Such a system may be regarded as falling within the category of predictive policing systems prohibited under the AI Act. Difficulties arise also in distinguishing between real-time and post-remote biometric identification, where the interval between data collection and individual identification is crucial for determining whether the AI system falls within a prohibited category or is classified as high-risk.

The criteria for determining risk levels remain contentious, particularly for systems that, while potentially impactful, do not meet the threshold for high-risk classification. As a result, such systems are not subject to mandatory compliance obligations, leaving a regulatory gap that may undermine the AI Act's capacity to adequately mitigate risks. For instance, place-based predictive policing – classified as posing only minimal risk under the AI Act, as discussed in Chapter 3 – illustrates the limitations of the current risk framework. As noted at the outset, even apparently simple location-based predictive tools can disproportionately impact certain communities by entrenching patterns of over-policing. This risk warrants subjecting such tools to the same regulatory requirements applied to systems classified as high-risk.

Another example is the French legislation adopted for the use of augmented cameras during the Olympic Games⁵. While such systems qualify as AI systems under the AI Act, they do not involve biometric identification, and whether they involve classification is debatable. Accordingly, they should not be categorized as either prohibited AI practices or high-risk systems, at least where they do not classify individuals on the basis of sensitive or protected attributes or characteristics. Nonetheless, the French law authorising their deployment introduces its own safeguards, which in some respects anticipates many of the requirements laid down by the AI Act for high-risk AI systems. A similar provision may be adopted also in different scenarios, such as in relation to systems used for place-based predictive policing.

⁵ Law No. 2023-380 of 19 May 2023, Article 10. Available at: https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000047561989.

CONCLUDING REMARKS

Even when the appropriate risk category can be identified without significant difficulty and the system is duly classified, doubts remain regarding the effective implementation of the requirements and obligations set forth in the AI Act. In Chapter 4, for example, I examined two of these requirements in detail – transparency and fairness – illustrating the significant challenges involved in applying them effectively, especially in the law enforcement context⁶. In particular, the high-risk regime relies heavily on technical standards, providers’ compliance and on national supervisory authorities’ capacity to enforce the rules. Therefore, whilst the AI Act constitutes a substantive regulatory framework that enhances the protection of individuals and public values, its efficacy is contingent upon robust implementation mechanisms, consistent national oversight, and rigorous enforcement by Member States.

Notably, as previously discussed, the AI Act establishes minimum standards that AI systems must meet in order to be placed on the market, put into service, or used within the European Union. Public administrations deploying AI systems retain, however, the discretion to impose more stringent requirements that exceed these baseline provisions. When developing AI systems in-house or procuring them through tendering procedures, they can introduce appropriate clauses and conditions requiring compliance with higher levels of data quality, transparency, and fairness. This possibility follows directly from the fact that public authorities – unlike private entities – are bound by specific principles such as transparency, fairness and non-discrimination, accountability, and the rule of law, which must remain fully respected even when public functions are carried out with the support of AI systems. For instance, with regard to transparency – albeit addressing only one of its many dimensions – the Italian legislator requires that public administrations procuring AI systems ensure the accessibility of the source code. This requirement ensures that the public authority retains the capacity to intervene in the system when necessary, including through engagement with the provider, thereby preventing situations in which the administration becomes dependent on opaque or inaccessible technological solutions⁷.

Regarding the latter question, clear boundaries are lacking in relation to the notion of “law enforcement”. Defining this concept is essential, given that the AI Act introduces several sector-specific provisions applicable to AI systems deployed in law enforcement contexts. Indeed, the AI

⁶ The situation may also shift in light of the European Commission’s recently proposed changes to digital legislation. The proposed ‘Digital Omnibus’ package also intervenes in the regulatory framework for AI systems, proposing targeted simplification measures to ensure timely, smooth, and proportionate implementation of certain of the AI Act’s provisions. See European Commission, *Proposal for a Regulation of the European Parliament and of the Council amending Regulations (EU) 2016/679, (EU) 2018/1724, (EU) 2018/1725, (EU) 2023/2854 and Directives 2002/58/EC, (EU) 2022/2555 and (EU) 2022/2557 as regards the simplification of the digital legislative framework, and repealing Regulations (EU) 2018/1807, (EU) 2019/1150, (EU) 2022/868, and Directive (EU) 2019/1024 (Digital Omnibus)*, Brussels, 19.11.2025, COM(2025) 836 final.

⁷ Italian Code of Public Procurement, Article 30.

CONCLUDING REMARKS

Act explicitly recognises this domain as particularly sensitive, given the profound implications that AI-supported decisions may have for fundamental rights and due process⁸.

In general, no difficulties arise when the system is employed for the prevention, investigation, detection, or prosecution of criminal offences; in such cases, the link to traditional policing functions is straightforward. As discussed in Chapter 1, there is, broadly speaking, a relatively consistent approach among Member States regarding the definitions of “criminal offences” and “criminal penalties”. By contrast, significantly greater uncertainty emerges when AI systems are deployed for the safeguarding or prevention of threats to public security, a category that remains both conceptually and operationally ambiguous. Despite this ambiguity, the notion of safeguarding public security has been invoked, for example, to justify the use of augmented or “intelligent” video surveillance during the French Olympic Games, and more generally to support the deployment of face recognition technologies and AI-enabled cameras at demonstrations, major sporting events, riots, and other large gatherings. These scenarios concern police activities aimed at maintaining law and order, crowd management, or the prevention of hypothetical risks, rather than the enforcement of criminal law strictly speaking. The use of AI to prevent threats to public security is therefore characterised by a fundamentally anticipatory logic. Unlike reactive policing, which is triggered by the commission of criminal offences, preventive measures engage police functions at an earlier stage and operate within a broader conceptual framework that extends beyond traditional criminal law categories.

Notwithstanding these complexities, establishing precise definitional boundaries for law enforcement remains a prerequisite for accurate risk classification within the AI Act’s regulatory architecture. For example, a real-time RBI system may be prohibited when used for law enforcement purposes, or it may be designated as high-risk when deployed in different contexts. Likewise, defining these boundaries accurately is crucial for determining whether an AI system falls within the high-risk category, since Annex III, point 6 explicitly refers to the law enforcement sector as one in which the use of AI poses particularly significant risks to fundamental rights.

More specifically, the AI Act introduces a specific regulatory framework for AI systems used in law enforcement, justified by the need to appropriately balance, on the one hand, the protection of fundamental rights and, on the other, the constitutional interests of public security and public order. Striking this balance is inherently delicate, and when AI technologies enter the law enforcement domain, the equilibrium risks being disrupted, potentially creating an imbalance that favours public authorities at the expense of individual rights⁹. For instance, the AI Act introduces specific provisions

⁸ AI Act, Recital 59.

⁹ This is expressly acknowledged by the AI Act. Recital 59 states that “[g]iven their role and responsibility, actions by law enforcement authorities involving certain uses of AI systems are characterised by a significant degree of power imbalance and may lead to surveillance, arrest or deprivation of a natural person’s liberty as well as other adverse impacts on fundamental rights guaranteed in the Charter”.

CONCLUDING REMARKS

in relation to transparency, as discussed in Chapter 4, acknowledging that certain disclosure obligations – such as informing individuals that they are interacting with an AI system or providing detailed explanations of its functioning – may be restricted where necessary to preserve the effectiveness of police operations or to protect public security. As a result, individuals subject to AI-assisted policing measures may receive less information about the system’s deployment, capabilities, or reasoning processes than they would in non-law-enforcement contexts. This built-in asymmetry reflects a broader tension: while law enforcement is treated as a high-risk domain warranting enhanced safeguards, operational secrecy and the need to prevent circumvention often justify limitations on transparency, potentially undermining accountability and the ability of affected individuals to meaningfully exercise their rights.

Furthermore, the safeguarding and prevention of threats to public security, which fall within the notion of law enforcement, shall be distinguished by the prevention of threats to national security, which lie outside the law enforcement sphere. As I discussed in Chapter 1, while in some situations this distinction is relatively clear, in others it becomes far more difficult to draw a precise boundary. This is particularly true in areas such as counter-terrorism, violent extremism, or the prevention of other serious crimes that may endanger the integrity of the State. In such contexts, activities aimed at safeguarding public security may overlap significantly with measures intended to protect national security, making it challenging to determine which regulatory framework applies. This ambiguity is not merely theoretical, rather it has practical implications for the applicability of the AI Act and thus the safeguards available to individuals. Indeed, AI systems that are developed, placed on the market, put into service, or used exclusively for military, defence, or national security purposes are excluded from the scope of the Regulation¹⁰. This exclusion, however, is strictly purpose-based: if an AI system originally intended for military, defence, or national security use is – either temporarily or permanently – employed for other purposes, such as civilian, humanitarian, law enforcement, or public security activities, the system is no longer covered by the exemption. In such cases, the system would fall within the scope of the Regulation, and its use would be subject to the corresponding obligations and compliance requirements.

The exclusion of national security from the scope of the AI Act remains one of the most contested features. While this carve-out is formally justified by the limits of EU competence, national security being the exclusive responsibility of the Member States¹¹, it nonetheless raises significant regulatory concerns. AI systems deployed in intelligence and defence contexts often pose substantial risks to

¹⁰ AI Act, Recital 24. “*As regards national security purposes, the exclusion is justified both by the fact that national security remains the sole responsibility of Member States in accordance with Article 4(2) TEU and by the specific nature and operational needs of national security activities and specific national rules applicable to those activities*”.

¹¹ TFEU, Article 4.

CONCLUDING REMARKS

fundamental rights, yet the exclusion prevents the AI Act from addressing these high-impact uses. Moreover, the exemption's purpose-based nature creates the potential for strategic reclassification of activities as national security to circumvent regulatory obligations, thereby generating uncertainty and possible fragmentation across Member States, each of which may adopt its own interpretation of what constitutes national security. Although the AI Act attempts to mitigate these risks by bringing systems within its scope whenever they are used for civilian, humanitarian, law enforcement, or public security purposes, the breadth and fluidity of the national security concept may still leave considerable gaps in oversight. As a result, the exclusion, while legally necessary, risks weakening the overall protective framework of the Regulation in precisely those domains where the intrusive potential of AI is most acute.

2. The delicate relationship between AI and personal data

Answering the question of whether the AI Act sufficiently protects individuals' fundamental rights is inherently limiting. As noted earlier, the AI Act does not operate in isolation. When AI systems are developed and deployed, additional regulatory frameworks become applicable. In particular, data protection legislation remains relevant, as it governs all situations involving the processing of personal data.

Before the entry into force of the AI Act, data protection legislation constituted the primary constraint on the development and deployment of AI systems. The GDPR and the LED imposed strict requirements on the collection, use, and retention of personal data, which often acted as a *de facto* barrier to certain AI practices, particularly those involving large-scale data processing or intrusive forms of profiling. In the absence of a dedicated regulatory framework for AI, therefore, data protection rules effectively served as the main legal mechanism for limiting or shaping the permissible uses of AI technologies within the European Union¹². Following the entry into force of the AI Act, data protection legislation has remained a crucial point of reference, as AI systems inherently rely on data, often including personal data, and it both complements and reinforces the framework established by the AI Act¹³.

¹² In May 2024, the EDPB published a noteworthy opinion regarding the use of facial recognition to facilitate airport passenger flows, permitting such processing only within the boundaries defined in the Opinion. EDPB, *Opinion 11/2024 on the use of facial recognition to streamline airport passengers' flow (compatibility with Articles 5(1)(e) and(f), 25 and 32 GDPR)*, 24 May 2024. Available at: https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-112024-use-facial-recognition-streamline_en.

¹³ On this topic see European Data Protection Board, *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models*, 17 December 2024. Available at: https://www.edpb.europa.eu/system/files/2024-12/edpb_opinion_202428_ai-models_en.pdf. The Italian Data Protection Authority stated that, while the GDPR focuses on the processing of (personal) data, AI regulation concerns the technology

CONCLUDING REMARKS

This close relationship emerges specifically in relation to the processing of personal data in the context of the development¹⁴ and deployment¹⁵ of AI models. This occurs whenever personal data is included in the datasets used, whether for training, validation, or testing of the model¹⁶. As discussed in Chapter 1, an AI system relies on an underlying AI model to achieve its intended purpose, integrating that model within a broader operational framework¹⁷. Similarly, personal data may be used as operational data for AI systems. For instance, an AI system for person-based predictive policing might use an AI model trained on historical personal, criminal, and social data to assess the risk of an individual engaging in future criminal activity. In particular, in the context of person-based predictive policing, the processing of personal data is not merely incidental but intrinsic to the operation of such AI systems. Nevertheless, incidental processing of personal data may also occur in the context of place-based predictive policing, depending on the data sources employed. Furthermore, predictive-policing systems may process personal data relating to criminal convictions and offences, which receive special protection under data-protection legislation¹⁸. Similarly, when AI systems are used in real time to identify individuals in publicly accessible spaces – for instance by matching their faces against a database – this entails the processing of personal data and, in particular, of special categories of personal data (*i.e.* biometric data).

With regard to the applicability of data protection rules¹⁹, any model that processes personal data falls squarely within the scope of data protection legislation and must adhere to all relevant principles and

used to carry out such processing and, therefore, the two are formally complementary *Intelligenza artificiale e ruolo della protezione dei dati personali. L'analisi di Ginevra Cerrina Feroni (Garante Privacy). Speech (AKey4biz, 14 febbraio 2023)*. For a discussion on the relationship between the two see P. FALLETTA, A. MARSANO, *Intelligenza artificiale e protezione dei dati personali: il rapporto tra Regolamento europeo sull'intelligenza artificiale e GDPR*, in *Rivista italiana di informatica e diritto*, 2024, 6(1), 120-137. See also the interesting volume F. PIZZETTI (Ed.), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Giappichelli, 2018.

¹⁴ “The development of an AI model covers all stages before any deployment of the AI model, and includes, inter alia, code development, collection of training personal data, pre-processing of training personal data, and training”. EDPB, Opinion No. 28/2024, cited, paragraph 18.

¹⁵ “The deployment of an AI model covers all stages relating to the use of an AI model and may include any operations conducted after the development phase”. EDPB, Opinion No. 28/2024, cited, paragraph 18.

¹⁶ Many changes will result from the Digital Omnibus Act mentioned above. In relation to data protection legislation “the amendments contained in this proposal aim to address those concerns notably by clarifying certain key definitions, for instance the notions of personal data; by facilitating compliance, for instance by supporting controllers with respect to the criteria and means to determine whether data resulting from pseudonymisation does not constitute personal data, in relation to information requirements and data breach notifications to supervisory authorities; as well as by clarifying certain aspects as to the processing of data for AI training and development”.

¹⁷ For the difference between AI systems and AI models see AI Act, Recital 97.

¹⁸ GDPR, Article 10.

¹⁹ The first step is to determine whether an AI model can be considered “anonymous”. With respect to the question of the anonymity of AI models, the European Data Protection Board concluded that the determination of whether an AI model is anonymous should be assessed on a case-by-case basis. An AI model may be deemed anonymous where there is sufficient evidence demonstrating that, through means reasonably likely to be employed: (i) personal data incorporated into the training process cannot be extracted from the model, and (ii) any outputs generated by the model cannot be associated with the data subjects whose personal data was used for training. In assessing whether these criteria are satisfied, the following elements shall be taken into account: the anonymisation techniques implemented in relation to the training data; all means reasonably likely to be used by the controller or by any other person to re-identify data subjects; and the controller’s evaluation of the risk of re-identification, both by itself and by various categories of third parties. In

CONCLUDING REMARKS

obligations. Therefore, as discussed in Chapter 1, when personal data are processed for law enforcement purposes, the processing must comply with the LED, which operates as a *lex specialis* in relation to the GDPR. Accordingly, law enforcement authorities remain fully bound to comply with the relevant obligations under data protection legislation when functioning simultaneously as data controllers and AI providers or deployers. The obligations arising under the AI Act therefore operate alongside, rather than in place of, the safeguards set out in instruments such as the LED and its national transpositions.

The intersection of data protection law and AI presents inherent complexities²⁰, with AI systems posing multifaceted risks to fundamental data protection principles. AI systems – especially those relying on large-scale, high-dimensional datasets – often depend on extensive data collection, prolonged retention periods, and the repurposing of data for model training, optimisation, and validation. These operational requirements sit uneasily with core LED principles, particularly data

cases where information relating to identified or identifiable individuals whose personal data were used to train the model can be obtained from the AI model through means that are reasonably likely to be used, it can be concluded that the model cannot be regarded as anonymous. To this end, controllers should adopt appropriate measures during the development phase to prevent or minimise the collection of personal data used for training, to reduce the identifiability of such data, to prevent their extraction, or to ensure that the model is resistant to attacks in accordance with the state of the art. Appropriate measures shall be taken also during the deployment phase in order to reduce the likelihood of obtaining personal data from an AI model. See EDPB, Opinion No. 28/2024, cited. Working Party 29, *Opinion 05/2014 on Anonymisation Techniques*, 10 April 2014. To this end, the data may be considered anonymous if it is not possible to single out, link and infer information from the supposedly anonymous dataset. WP29 Opinion 05/2014, cited, 24. The leading case was CJEU judgment of 19 October 2016, Case C-582/14, *Breyer v Bundesrepublik Deutschland*, ECLI:EU:C:2016:779. With the recent CJEU judgment of 4 September 2025, Case C-413/23, *European Data Protection Supervisor v Single Resolution Board*, ECLI:EU:C:2025:645. The Court held that pseudonymized data may, depending on the circumstances assessed on a case-by-case basis, effectively constitute anonymous data for the recipient, provided that the recipient does not have reasonable means (such as costs, time, or resources) to re-identify the data subject, that is, the individual to whom the data relates. In relation to re-identification, the EDPB specifies that the following aspects shall be considered: “a. the characteristics of the training data itself, the AI model and the training procedure; b. the context in which the AI model is released and/or processed; c. the additional information that would allow the identification and may be available to the given person; d. the costs and amount of time that the person would need to obtain such additional information (in case it is not already available to them); and e. the available technology at the time of the processing, as well as technological developments”. EDPB, Opinion 28/2024, cited, paragraph 41. In the Opinion No. 28/2024, the EDPB provided a non-prescriptive and non-exhaustive list of possible elements that may be considered by DPAs when assessing a controller’s claim of anonymity: selection of sources used to train the AI model, preparation of data for the training phase (e.g. pseudonymisation measures, reasons for not using them, data minimisation strategies), methodological choices regarding the training (e.g. differential privacy), measures regarding outputs of the model (i.e. lowering the likelihood of obtaining personal data related to training data from queries), document-based audits that include an evaluation of the chosen measures and of their impact to limit the likelihood of identification, AI model testing and resistance to attacks, documentation of the processing operations including all information identified in EDPB, Opinion 28/2024, cited, paragraph 58. Existing research identifies vulnerabilities in AI models that may lead to the processing or disclosure of personal data, either during model deployment (e.g. via APIs or prompt interfaces) or through targeted attacks. Ongoing research, particularly on training data extraction, demonstrates that personal data can sometimes be intentionally or inadvertently retrieved from AI models. See M. VEALE, R. BINNS, L. EDWARDS, *Algorithms that remember: model inversion attacks and data protection law*, in *Philosophical Transactions of the Royal Society*, 376(2133), 2018, Article 20180083, 1-16. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3212755. N. Carlini, F. Tramèr, E. Wallace, M. Jagielski, A. Herbert-Voss, K. Lee, A. Roberts, T. Brown, D. Song, U. Erlingsson, A. Oprea, C. Raffel, *Extracting Training Data from Large Language Models*, 15 June 2021. Available at: <https://arxiv.org/pdf/2012.07805>.

²⁰ A fundamental contribution on this topic is G. SARTOR, F. LAGIOIA, *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*, 2020. G. FINOCCHIARO, *Intelligenza Artificiale e protezione dei dati personali*, in *Giurisprudenza italiana*, 2019, 1673-1674.

CONCLUDING REMARKS

minimisation, which mandates that personal data be adequate, relevant, and limited to what is strictly necessary for the specific law enforcement purpose, and the obligation to restrict retention periods to what is objectively required. Furthermore, AI systems designed to learn from historical data may implicitly incentivise broader and longer-term data storage, creating structural tensions with these principles. Simultaneously, certain data protection principles find counterparts within the AI Act. The principle of fairness, examined in Chapter 4, exemplifies this convergence. In data protection legislation, fairness seeks to address power asymmetries between data controllers – namely, law enforcement authorities – and data subjects, primarily by requiring controllers to identify, analyse, and mitigate bias within the datasets.

Data protection legislation establishes, as a fundamental and overarching principle, that any processing of personal data must be grounded in a legal basis²¹. This requirement serves as the foundation for ensuring that personal data are handled lawfully. Without a clearly defined legal basis, the processing is considered unlawful, regardless of the purposes pursued or the safeguards in place. In recent years, DPAs and courts have repeatedly intervened to stop the processing of personal data – particularly automated processing – when an appropriate legal basis was absent²², including in the law enforcement context. As discussed in Chapter 2, the Italian DPA has repeatedly suspended AI projects and solutions developed at both the national and local levels due to the absence of an adequate legal basis. This includes the SARI Real-Time system, as well as initiatives implemented by the municipalities of Como, Lecce, Arezzo, and Turin. Similar conclusions have been reached by other European DPAs – such as in the case concerning the Länder Hesse and Hamburg – and by courts across Europe, including in the United Kingdom.

However, the mere existence of a legal basis is not sufficient. The legal basis must also meet certain qualitative standards that ensure legal certainty and protect individuals' rights. More specifically, it must clearly and specifically define the type of processing that is authorised, the purposes it serves, and the limits within which it may operate. This includes outlining the categories of data that may be processed, the actors involved, the safeguards to be applied, and the conditions under which the processing may be carried out. The central importance of a robust legal basis is clearly illustrated by the German Federal Constitutional Court's judgment of January 2023, which declared unconstitutional the statutory provisions underpinning the use of the hessenDATA platform acquired by the Land of Hesse from the U.S. company Palantir²³. The Court found that the legal framework

²¹ LED, Articles 4 and 8; GDPR, Articles 5 and 6.

²² M. VAN BEKKUM, F.Z. BORGESIU, *Digital welfare fraud detection and the Dutch SyRI judgment*, in *European Journal of Social Security*, 23(4), 2021, 323-340. <https://doi.org/10.1177/13882627211031257>.

²³ *Bundersverfassungsgericht*, I Senate (President Stephan Harbarth), Judgement 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20. Available at: https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2023/02/rs20230216_1bvr154719en.html.

CONCLUDING REMARKS

failed to provide sufficiently precise and proportionate safeguards governing the collection and analysis of personal data through such an AI-enabled investigative tool. Therefore, where public authorities seek to deploy advanced technological systems for legitimate objectives, such as security or crime prevention, the underlying legal basis must meet stringent standards of clarity, specificity, and proportionality in order to comply with constitutional and data protection requirements.

Any discussion concerning the legal basis for the processing of personal data must necessarily take into account the AI Act's provisions. First and foremost, the AI Act's prohibitions on certain AI practices necessarily limit the situations in which personal data can be lawfully processed under European data protection legislation. Indeed, insofar as personal data are processed, they must always be processed lawfully. Consequently, where the AI Act expressly prohibits the development and/or deployment of a particular AI practice, any related processing of personal data must likewise be considered unlawful, irrespective of whether it would otherwise meet the principles and conditions established under the applicable data protection framework. In this respect, the AI Act and data protection legislation operate in a relationship of normative complementarity²⁴. Second, as a general rule, the AI Act does not create a new legal basis for processing personal data within AI systems, nor does it replace the GDPR or the LED requirements for establishing a lawful basis for such processing²⁵. However, the AI Act does set out a few limited exceptions in which personal data may be processed under its framework. One key exception concerns the processing of personal data within AI regulatory sandboxes – controlled experimental or testing environments – provided that specific conditions are met²⁶. Another relates to the processing of personal data in the context of RBI systems. As discussed in Chapter 3, because the use of RBI systems entails significant risks, the AI Act

For an initial comment on the judgment see R. ORLANDI, *Uso poliziesco dell'intelligenza artificiale. L'insegnamento del Bundersverfassungsgericht*, in *Cassazione Penale*, 7/8, 2023, 2167-2180; M. MERLER, *La Corte costituzionale tedesca si pronuncia sul trattamento automatizzato dei dati in contesti di polizia*, in *Giornale di diritto amministrativo*, 2023, 5, 685-694.

²⁴ For a discussion on the protection of personal data in law enforcement contexts see Consultative Committee Of The Convention For The Protection Of Individuals With Regard To Automatic Processing Of Personal Data, *Practical guide on the use of personal data in the police sector*, Strasbourg, 15 February 2018.

²⁵ For instance, in Italy, the recently approved Law No. 132/2025 (Legge 23 settembre 2025, n. 132 Disposizioni e deleghe al Governo in materia di intelligenza artificiale. Available at: [25 and the Minister of the Interior's Decree 24 May 2017.](https://www.gazzettaufficiale.it/eli/id/2025/09/25/25G00143/sg.)

²⁶ The AI Act, and Article 59 in particular, provides “the legal basis for the providers and prospective providers in the AI regulatory sandbox to use personal data collected for other purposes for developing certain AI systems in the public interest within the AI regulatory sandbox, only under specified conditions, in accordance with Article 6(4) and Article 9(2), point (g), of Regulation (EU) 2016/679, and Articles 5, 6 and 10 of Regulation (EU) 2018/1725, and without prejudice to Article 4(2) and Article 10 of Directive (EU) 2016/680. In particular, this Regulation should not provide a legal basis in the meaning of Article 22(2), point (b) of Regulation (EU) 2016/679 and Article 24(2), point (b) of Regulation (EU) 2018/1725”. AI Act, Recital 140.

CONCLUDING REMARKS

introduces specific rules governing their deployment. In particular, the AI Act establishes a general prohibition on the use of real-time RBI systems, allowing them only under narrowly defined exceptions. Accordingly, a Member State may permit such use within its territory only if it has expressly provided for the possibility of authorising it in detailed provisions of its national law. Because these rules constitute *lex specialis*²⁷, these provisions override and take precedence over data protection rules enshrined in Article 10 of the LED – which deals with the processing of special categories of personal data – whenever real-time remote biometric identification in publicly accessible spaces for law enforcement purposes is involved²⁸. Therefore, for this specific use, the AI Act fully regulates not only the use of the technology, but also the biometric data processing involved. In any case, the AI Act does not provide the necessary legal basis for the processing of personal data required by Article 8 of the LED²⁹. Thus, the AI Act regulates *whether* and *under what* conditions authorities may use such AI systems, but it does not, on its own, authorise the underlying personal data processing. A further exception concerns the processing of special categories of personal data for the purpose of detecting biases, as discussed in Chapter 4. Acknowledging that discriminatory outcomes are not limited to high-risk AI systems, the European Commission has recently proposed expanding the legal basis for processing such data to detect and correct bias, subject to certain safeguards³⁰.

By contrast, in some cases no processing of personal data is involved. It is conceivable that certain place-based systems operate without relying on personal data and therefore fall outside the scope of application of the LED. This appears to be the case for KeyCrime, used by the municipality of Milan, and the XLAW software employed by the municipality of Naples, neither of which processes personal data. As a result, their use does not require a specific legal basis for data processing.

In conclusion, the LED and its national transpositions complement the AI Act by providing additional layers of protection for individuals. Nevertheless, effective coordination between these instruments is essential, as the LED introduces obligations that intersect with those established under the AI Act. This is evident, for example, in the relationship between the DPIA, required under the LED, and the FRIA, mandated by the AI Act, which must be carefully aligned. A similar need for coordination arises with respect to data quality: while the LED imposes general obligations in this regard, the AI

²⁷ AI Act, Recital 38.

²⁸ Where real-time biometric identification systems are deployed for purposes unrelated to law enforcement, the AI Act's strict *lex specialis* provisions are inapplicable. Accordingly, Article 10 of the LED remains the applicable framework for biometric processing by competent authorities in cases that do not involve real-time operation or occur outside public spaces.

²⁹ AI Act, Recital 38.

³⁰ European Commission, *Proposal for a Regulation Of The European Parliament And Of The Council amending Regulations (EU) 2024/1689 and (EU) 2018/1139 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI)*, Brussels, 19.11.2025, COM(2025) 836 final, Recital 6.

CONCLUDING REMARKS

Act elaborates more detailed and system-specific requirements. Coordination is also crucial when considering Article 11 of the LED on fully automated decision-making and Articles 14 and 26 of the AI Act, which effectively restricts such scenarios by requiring that high-risk AI systems operate under meaningful human oversight. Despite these points of convergence, it must be noted that the relationship between the two instruments remains contentious, particularly given the tensions that AI technologies create for certain data protection principles, such as purpose limitation and storage limitation. Moreover, the LED itself presents specific challenges. As a directive, it has been transposed differently across Member States, resulting in variations and lingering uncertainties regarding its practical implementation. These divergences complicate the overall interaction between the LED and the AI Act and may affect the consistency of protections afforded to individuals throughout the European Union.

3. Insights for human rights law

Human rights law and jurisprudence may offer interesting insights in delineating the limits of state authority and identifying the safeguards necessary to ensure the effective protection of individuals' fundamental rights. A relevant judgment in this regard is the well-known case *Glukhin v. Russia*, delivered by the ECtHR in July 2023³¹, discussed in Chapter 2. The case concerned the use by the Russian authorities of CCTV cameras equipped with face recognition technology, installed in public spaces, which enabled them to identify a protestor who was subsequently convicted in administrative proceedings for failing to notify the authorities of his intention to hold a solo demonstration. The applicant alleged a violation of Article 8 of the ECHR, on the ground that the authorities had unlawfully processed his personal data, as well as a violation of Article 10 of the ECHR, arguing that the use of face recognition-enabled CCTV had unlawfully interfered with his right to freedom of assembly³². Human rights law and the corresponding jurisprudence are structured around the well-established three-part test of legality, necessity, and proportionality, a framework consistently recognised within the European human rights system³³. Analogously, Article 52 CFREU provides that any restriction on the rights guaranteed therein – including the right to the protection of personal data – must be prescribed by law. This requires, in particular, that the legal basis enabling an

³¹ ECtHR, *Glukhin v. Russia*, cited.

³² With regard to the specific matter of new technologies and freedom of assembly see United Nations High Commissioner for Human Rights, *Impact of new technologies on the promotion and protection of human rights in the context of assemblies, including peaceful protests*, UN Doc. A/HRC/44/24, report, 24 June 2020.

³³ Council of Europe, *Recommendation No. R (87) 15 of the Committee of Ministers regulating the use of personal data in the police sector*, 17 September 1987; Consultative Committee of the Convention for the protection of Individuals with regard to Automatic Processing of Personal Data (ETS 108), *Guidelines on Facial Recognition*, 2021; EDPB, *Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement*, 26 April 2023;

CONCLUDING REMARKS

interference with fundamental rights must itself define with sufficient clarity and precision the scope, conditions, and limits of the restriction. Moreover, any limitation must respect the essence of the right at issue, genuinely pursue an objective of general interest recognised by the Union or the protection of the rights and freedoms of others, and comply with the principle of proportionality. These requirements must be interpreted in light of the case law of the CJEU³⁴.

The mere existence of a legal basis is thus not sufficient. The law authorising an interference must also meet the required standards of quality, meaning that it must be clear, precise and foreseeable in its application. It must further establish minimum safeguards to ensure that individuals whose personal data are processed benefit from adequate guarantees enabling them to effectively protect their data against the risk of misuse. These quality requirements become all the more essential where personal data are subject to automated processing, particularly in contexts where the available technologies are increasingly sophisticated³⁵. Therefore, legislation authorising interference must clearly define the scope of the interference; include minimum safeguards against abuse and arbitrariness; be predictable for individuals; comply with necessity and proportionality, limiting interference to what is strictly necessary.

Following the judgment of the Federal Constitutional Court, the Land Hesse amended the provision that continues to serve as the legal basis for the use of hessenDATA. In contrast to the earlier version – which the Court had declared unconstitutional – the revised provision represents an attempt to give proper effect to the Court’s reasoning and to align the statutory framework with the constitutional requirements it articulated³⁶. In particular, under the new provision the police may combine various categories of lawfully collected data (case files, police systems, traffic data, telecommunications data, seized items, police information exchange); targeted queries from state registers and select internet sources may also be included; limitation on the use of traffic and telecommunications data is identified when acting under the third category (serious planned offences). Among the safeguards enshrined in the provision an administrative regulation must ensure a limitation of the data volume, ensure proportionality, and protect uninvolved persons. To this end, personal data of uninvolved persons may not be included, data obtained from residential surveillance or online searches is

³⁴ CJEU, Judgement 24 February 2024, *SIA ‘SS’ v. Valsts ieņēmumu dienests*, Case C-175/20, ECLI:EU:C:2022:124. CJEU, Judgement 1st August 2022, *OT v. Vyriausioji tarnybinės etikos komisija*, Case C-184/20, ECLI:EU:C:2022:601.

³⁵ *Glukhin v. Russia*, paragraphs 75 and 77.

³⁶ HSOG, §25a, as amended after the declaration of unconstitutionality, held that police authorities may further process stored personal data using an automated data-analysis application: 1) where necessary to avert a danger to the existence or security of the Federation or a Land, or to the life, physical integrity or liberty of a person, or to property of significant value whose preservation is required in the public interest, or where equivalent environmental damage is expected (averting specific dangers); 2) where factual indications justify the assumption that, within a foreseeable period, criminal offences of significant importance will be committed in a manner at least specified in kind, and further processing is necessary to prevent these offences (averting concretised dangers); 3) where factual indications justify the assumption that serious or particularly serious offences are planned, and further processing is necessary to prevent these offences (preventive combating of crime). Paragraph 2 also clarifies the data the sources of data that may be processed.

CONCLUDING REMARKS

excluded, access to the system is subject to strict access controls and continuous logging, and every use must be justified. Moreover, police authorities must ensure that discriminatory algorithms are neither developed nor used.

The principles discussed above apply equally where data processing is characterised as merely preparatory to the development of new technologies, such as the training of AI models. This was precisely the position taken by the Italian DPA in relation to the MARVEL project, discussed in Chapter 2³⁷. In that case, the DPA found that no legal framework – adequate in terms of both normative rank and substantive quality – had been demonstrated to justify the processing of personal data carried out by a public body, namely the municipality acting as controller, within the scope of the MARVEL project. Consequently, the interference with the fundamental rights and freedoms of the individuals whose data were collected and processed lacked a lawful and sufficiently robust legal basis and was therefore discontinued.

By contrast, the French Conseil constitutionnel upheld the adequacy of the legal basis for the automated processing of data provided by Article 10 of Law No. 2023-380 of 19 May 2023, which authorised the use of automated video-surveillance systems during the Olympic Games³⁸. Bringing an end to a long-standing debate over the constitutional legitimacy of such systems, the Conseil constitutionnel found that the use of so-called “augmented cameras” is compatible with the French Constitution³⁹. More specifically, it held that the legislation had provided all safeguards necessary to protect the right to respect for private life. Indeed, it is the legislature that bears responsibility for defining the fundamental guarantees afforded to citizens in the exercise of their public liberties. Accordingly, it must ensure a proper balance between, on the one hand, the constitutionally recognised objective of preventing public order offences and, on the other, the right to privacy⁴⁰.

4. Organisational changes for responsible AI innovation

A final consideration for answering the research question stems from the recognition that the AI Act’s provisions are highly complex to implement and will therefore require organisational adjustments within law enforcement structures for an effective implementation. Although many obligations are

³⁷ Garante per la protezione dei dati personali, *Decision 11 January, No. 5*. Available at: <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9977020>.

³⁸ Law No. 2023-380 of 19 May 2023, Article 10. Available at: https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000047561989.

³⁹ Conseil Constitutionnel, *Décision No. 2023-850 DC du 17 mai 2023*. Available at: https://www.legifrance.gouv.fr/download/pdf?id=yUQ_DLw0skm8dWcA3FHSXGOhrN4FcEKuXTAQRsVaHPI, par. 26-49.

⁴⁰ Conseil Constitutionnel, *Décision No. 2023-850*, paragraph 32.

CONCLUDING REMARKS

directed primarily at providers – and, as noted, law enforcement authorities generally act as deployers – deployers themselves are nonetheless subject to a number of obligations that they must comply with.

Responsible AI innovation, in the present author's assessment, necessitates coordinated intervention across three distinct yet complementary dimensions⁴¹: an organisational culture that embraces innovation, remains attentive to changes in laws and regulations, and is prepared to adapt accordingly; a workforce with the expertise needed to make informed decisions about AI; and processes that support responsible AI innovation.

4.1. A proper organisational culture

The organisational culture of a law enforcement agency can be defined as the collective values, objectives, attitudes, and practices that characterise the institution⁴². Typically, these elements are formalized within a strategy, policy, or vision statement that conveys the agency's overarching aspirations and objectives⁴³. Developing responsible AI innovation requires, first and foremost, an organisational structure that consistently upholds key governance principles⁴⁴.

At the core of such a structure is a commitment to human agency and oversight, ensuring that people, rather than automated systems, remain in control of critical decisions. Equally important is technical robustness and safety, which calls for systems that are reliable, secure, and resilient to errors or malicious interference. A responsible approach to AI must also focus on the prevention of harm, safeguarding individuals and communities from physical, psychological, economic, or societal risks. Fairness, discussed in Chapter 4 Part II, plays a central role, demanding that AI systems avoid discrimination and promote equal treatment across different demographic groups. Alongside this, transparency is essential for enabling end-users to understand how AI systems function and to make informed decisions, while accountability ensures that clear lines of responsibility exist throughout the AI lifecycle. Finally, strong data protection practices are critical to uphold privacy rights⁴⁵. Although

⁴¹ For the following discussion specific reference is made to UNICRI, INTERPOL, *Responsible AI Innovation in law enforcement. AI toolkit. 03. Organizational Roadmap*, February 2024.

⁴² UNICRI, INTERPOL, *Organizational Roadmap*, cited, 5.

⁴³ S. CHOURASIA, A. DHAMA, G. BHARDWAJ, *AI-Driven Organizational Culture Evolution: A Critical Review, 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)*, Gautam Buddha Nagar, India, 2024, 1839-1844.

⁴⁴ In particular, it is the chief of police or the executive leadership that shall promote an organisational culture around responsible AI innovation, also through incentives of different kind.

⁴⁵ The Guidelines ensure that the development, deployment and use of AI systems meets seven key requirements for being considered “trustworthy”: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) environmental and societal well-being and (7) accountability. They also consider technical and non-technical methods to ensure the implementation of those requirements. Chapter III provides a concrete and non-exhaustive Trustworthy AI assessment list aimed at operationalising the key requirements set out in Chapter II.

CONCLUDING REMARKS

the AI Act refrains from expressly codifying a trustworthiness framework, the cumulative effect of its requirements is to construct precisely such an architecture. Similarly, the *Ethics Guidelines for Trustworthy AI*, referenced by the AI Act, serve as a valuable and influential benchmark for trustworthy AI development.

As recognised by the study conducted by UNICRI, the development of an organisational culture oriented toward responsible AI innovation is closely intertwined with the need to foster public trust⁴⁶. Such trust is not merely a desirable outcome, rather it is a prerequisite for law enforcement authorities, in any country, to fulfil their functions and duties effectively: when the public has confidence that AI systems are used ethically, transparently, and in a manner that respects fundamental rights, law enforcement authorities are better positioned to carry out their mandate, maintain legitimacy, and strengthen cooperation with the communities they serve. Despite its importance, this process is not straightforward. It can be lengthy, complex, and resource intensive. Considering the AI lifecycle described in Chapter 1, it should be evident that the decision to deploy an AI system must not be underestimated. As noted in that Chapter, the problem owner (e.g., the law enforcement authority) must first determine whether an AI system is truly the most appropriate and effective solution to the issue at hand. Accordingly, before deciding to implement AI a thorough self-assessment is required⁴⁷. Fostering an appropriate organisational culture also means promoting support for, and engagement with, a diverse range of stakeholders. Although achieving such engagement may appear particularly challenging in the law enforcement context, the UNICRI study identifies it as a guiding principle for responsible AI innovation⁴⁸. Engaging stakeholders not only enhances transparency and legitimacy but also ensures that different perspectives and concerns are taken into account throughout the AI development and deployment process. Stakeholders who should be involved include those who will directly use the system (such as police officers), regulators, relevant interest groups, and individuals or communities affected by the deployment of the AI system. All these actors should – each to an appropriate extent – participate in the process. Despite the transparency barriers discussed in Chapter 4 Part I, clear communication regarding the needs, objectives, and values underpinning the development, procurement, and use of an AI system in the law enforcement context remains essential⁴⁹. Such communication enhances transparency and fosters public trust by giving stakeholders a sense of inclusion in decisions that may affect their rights. Alongside external communication to the public, effective internal communication is equally crucial. Clearly explaining to law enforcement personnel the need for and added value of integrating AI into their operational

⁴⁶ UNICRI, INTERPOL, *Organizational Roadmap*, cited, 6.

⁴⁷ UNICRI, INTERPOL, *Organizational Roadmap*, cited, 6.

⁴⁸ UNICRI, INTERPOL, *Organizational Roadmap*, cited, 6-7.

⁴⁹ Otherwise, the introduction of AI may be viewed merely as an added burden, requiring additional steps, for example, the submission of extra documentation.

CONCLUDING REMARKS

work, including how it can enhance efficiency and effectiveness, is essential to prevent or at least minimise internal resistance⁵⁰.

4.2. People

People remain at the core of responsible AI innovation. As should be clear from the previous Chapters, the AI Act introduces a particularly complex regulatory framework that demands specialised knowledge and expertise, going well beyond the skills typically required for digital innovation more generally. As discussed in Chapters 3 and 4, deployers are among the primary addressees of obligations under the AI Act. These responsibilities encompass the continuous monitoring of an AI system's functionality after deployment, detecting and reporting errors or malfunctions, and taking appropriate action – such as correcting, suspending, recalibrating, or decommissioning the system – in the event of failure. Therefore, even when a public authority decides to rely on external expertise for the technical development of an AI system and procures such products from third parties, a degree of internal technical competence remains indispensable. Only with such expertise can deployers make informed and responsible decisions about the procurement, use, and oversight of AI systems.

The AI Act addresses this imperative through an AI literacy obligation enshrined in its general provisions⁵¹. This obligation is formally addressed to both providers and deployers, mandating that they ensure their staff and all individuals involved in the operation or use of AI systems possess adequate knowledge and skills⁵². From a broader perspective, European countries have so far invested insufficiently in AI-related training, particularly with regard to civil servants. This landscape, however, is beginning to shift. Italy offers an instructive example: Law No. 132/2025 has delegated the Government to adopt one or more legislative decrees aimed at aligning national legislation with Regulation (EU) 2024/1689, and in doing so has laid down a series of specific guiding principles to which the Government must adhere. These include the introduction of AI literacy and training programmes targeting the general public, as well as tailored training pathways to be organised by professional orders, representative trade associations, and aggregated professional bodies directed at practitioners and sector-specific operators. The delegation further mandates the strengthening of STEM competencies within school curricula with a view to encouraging students, through targeted

⁵⁰ For a discussion on police pushback see S. BRAYNE, *Predict and Surveil: Data, Discretion, and the Future of Policing*, OUP USA, 2021, 74. The primary source of resistance stems from concerns about change, as well as apprehension about being monitored or observed by the new technology.

⁵¹ AI Act, Article 4.

⁵² The European Commission's recently proposed amendments, however, would transform this obligation from one imposed on providers and deployers to one imposed on the Commission and Member States themselves.

CONCLUDING REMARKS

and personalised guidance, to pursue higher education in those fields. Finally, universities and higher technical institutes (ITS Academy) are required, consistently with their respective cultural and professional profiles, to incorporate training activities fostering both the technical understanding and the legally informed use of AI technologies, as well as the capacity to critically interpret the outputs generated by such systems. Italy, however, still lacks comprehensive AI training programmes for civil servants and public managers alike. Other jurisdictions have demonstrated considerably more substantive investment in building AI competencies within the public sector⁵³. A notable example is the United States AI Training Act, which mandates the Office of Management and Budget (OMB) to establish a dedicated AI training programme designed to ensure that the federal workforce possesses the knowledge and skills required to engage effectively with AI systems⁵⁴.

Law enforcement personnel are no exception. As end-users of AI systems, they must receive adequate training on the functioning, capabilities, and limitations of such systems, a prerequisite for ensuring their correct and responsible deployment. For law enforcement personnel, general AI literacy must be complemented by sector-specific expertise, given that the deployment of AI in policing raises distinct operational, ethical, and legal concerns. It is only through this deeper, context-sensitive understanding that officers can exercise sound judgement, resist over-reliance on automated outputs, maintain meaningful human oversight, and preserve accountability in a domain where decisions frequently carry significant consequences for the individuals concerned.

Apart from specialised technical competencies, a wide range of additional competencies is required⁵⁵. Above all, domain expertise, particularly across the various areas of policing, remains indispensable to the sound development and deployment of AI systems in this field. Broad knowledge of relevant laws, regulations, policies, and procedures is equally necessary. Moreover, responsible AI innovation often requires the involvement of professionals from diverse backgrounds, including ethicists, human rights specialists, and other experts who can help ensure that the technology is developed and deployed in a manner consistent with legal, ethical, and societal expectations. As discussed in relation to the Fundamental Rights Impact Assessment (FRIA), for instance, the participation of human rights experts is crucial, as their input significantly enhances the quality, robustness, and credibility of the

⁵³ MARCHETTI, TORCHIA, *AI and Public Administration*, cited.

⁵⁴ Available at: <https://www.congress.gov/bill/117th-congress/senate-bill/2551>.

⁵⁵ The UNICRI's study identifies four different competencies: technical competencies, consisting of the knowledge and skillset required to apply technical methods to solve a particular problem; domain competencies, referring to competencies around policing, its role in society and the criminal justice system; governance competencies, including competencies around human rights law and the ethical frameworks required to manage and implement AI systems responsibly; socio-cultural competencies, referring to the competencies required to manage an agency's public engagement and build public trust with awareness of the social and cultural context.

CONCLUDING REMARKS

assessment⁵⁶. Similar considerations apply to ethicists, given the centrality of ethical principles to the development and deployment of AI systems, as explored in Chapter 3.

Since a public authority acting as a deployer may not have all these profiles within its organisational structure, it may be necessary to establish partnerships with academia, industry, or civil society. Such collaborations can provide access to specialised knowledge and help ensure that AI systems are assessed, implemented, and monitored in a manner consistent with fundamental rights and ethical standards⁵⁷. As a concluding remark, I firmly believe that assembling teams that reflect diversity in gender, age, ethnicity, disability, and other characteristics is a crucial first step, also in reducing the risk of discrimination arising during the development of AI systems⁵⁸.

4.3. Processes

Introducing AI systems into the law enforcement context inevitably entails a range of risks. In this Thesis, the focus has primarily been on those applications that pose some of the highest risks to fundamental rights and freedoms, particularly predictive policing and biometric identification, discussed in Chapter 2. However, these practices do not represent the only uses of AI within the sector. Other applications, such as text and speech analysis⁵⁹, as well as AI systems for process optimisation and workflow automation are already widely deployed⁶⁰. In contrast to predictive policing and biometric identification, however, these latter applications generally appear to involve lower levels of risk. Nonetheless, responsible AI innovation requires, first and foremost, a clear awareness of the specific risks associated with each type of AI system. Only through such awareness can appropriate mitigation strategies be designed and implemented. In this context, therefore, “processes” can be understood as the concrete initiatives, actions, and activities undertaken by the organisation to develop and deploy AI systems in a responsible manner⁶¹.

While the AI Act primarily focuses on mitigation strategies to be implemented by providers – which, as noted several times, is generally not the role played by law enforcement authorities, who typically act as deployers – it is nonetheless crucial that deployers also establish internal processes to meet

⁵⁶ A. MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template*, in *Computer Law & Security Review*, 2024, 1-18. Available at: <https://www.sciencedirect.com/science/article/pii/S0267364924000864>.

⁵⁷ UNICRI, INTERPOL, *Organizational Roadmap*, cited, 12.

⁵⁸ UNICRI, INTERPOL, *Responsible AI Innovation in law enforcement. AI toolkit. 02. Principles for Responsible AI Innovation*, February 2024, 24.

⁵⁹ UNICRI, INTERPOL, *Responsible AI Innovation in law enforcement. AI toolkit. 01. Introduction to Responsible AI Innovation*, February 2024, 9-10.

⁶⁰ UNICRI, INTERPOL, *Introduction to Responsible AI Innovation*, cited, 11-12.

⁶¹ UNICRI, INTERPOL, *Organizational Roadmap*, cited, 20.

CONCLUDING REMARKS

their obligations under the AI Act and to uphold the ethical principles discussed above. In this final section, I have identified several specific processes that should be followed to support responsible AI innovation. These processes are not intended to be exhaustive and may serve as a basis for further research and exploration.

First and foremost, the law enforcement authority – identified as the problem owner in the AI lifecycle described in Chapter 1 – should conduct a thorough assessment of its needs, existing capabilities, and the potential role of AI systems within its organisation. This assessment must, first and foremost, determine whether developing, acquiring, or integrating an AI system is necessary and appropriate, and identify the requirements for doing so. It should evaluate the nature of the problem to be addressed, the potential added value of AI, the associated risks, and whether alternative non-AI solutions may be more suitable. Only if the assessment concludes that an AI system offers the most appropriate and proportionate solution should the problem owner proceed to the next stage. For instance, in relation to the adoption of hessenDATA, the Land Government of Hesse demonstrated that the growing use of digital media and communication technologies – particularly in the context of terrorist and extremist violence as well as organised and serious crime – has resulted in police authorities being confronted with ever-increasing volumes of data. These data streams are not only larger but also more heterogeneous in terms of quality, structure, and format. According to the Land Government, automated data analysis has thus become essential for effective police action, as critical information related to such crimes is often difficult, if not impossible, to retrieve through conventional searches of police records, especially when decisions must be made within tight timeframes⁶². In such circumstances, the use of AI systems can therefore be regarded as a suitable and proportionate solution to achieve the intended purpose⁶³.

Once the public authority has determined that AI is the most suitable solution, it can proceed to define the problem more precisely and identify the relevant characteristics, requirements, and constraints of the AI system to be implemented. A critical aspect of responsible AI innovation is ensuring that developers remain in close contact with the problem owner – namely, the law enforcement authority – so that the development of the AI system remains aligned with the organisation's values and objectives.

One of the major challenges at this stage is that public authorities are not always in a position to effectively negotiate these aspects, particularly when dealing with large international companies that generally possess greater technical expertise and bargaining power. This dynamic was also visible in

⁶² *Bundersverfassungsgericht*, cited, paragraph 52.

⁶³ Given the importance that the legislator may attribute to the powers in question for the performance of state functions and police work in Hesse, the Constitutional Court considered a temporary continuation of the provision preferable to declaring it void.

CONCLUDING REMARKS

the case of hessenDATA, where the counterpart was the big U.S. company Palantir. The company offered its system largely on the basis of its existing product design, leaving limited room for adaptation to the evolving requirements and safeguards later reflected in the AI Act. As a result, several key regulatory and ethical considerations were not sufficiently integrated at the outset, illustrating the difficulty public authorities may face when procuring complex AI tools from powerful multinational vendors. While the operating system responded effectively to the needs of the Hessian police, the same cannot be said for the Norwegian police, who announced the termination of their project in 2020, before it ever became operational. In this case, significant tensions emerged between Palantir's IT consultants and the Norwegian Data Protection Officers (DPOs). They held fundamentally divergent views on the project, particularly regarding data protection requirements, the system design, and compliance with national and European privacy standards. The resulting disagreements ultimately prevented the project from progressing and underscored the importance of aligning technological ambitions with robust data protection safeguards from the outset.

The Norwegian case also highlights a core principle of responsible AI innovation: the ability to pause or recalibrate an AI system when necessary⁶⁴. Recognising when an AI system no longer delivers value, no longer aligns with the organisation's objectives and values, or when the circumstances that originally justified its deployment have changed, is essential. Similarly, substantial public pushback or failure to meet predefined performance metrics should trigger a reassessment. In such situations, thus, the agency must be prepared to pause, revise, or even discontinue the system. As Sarah Brayne observes, law enforcement authorities' reluctance to push back against data-analysis technologies can also be explained by political and institutional factors. Developing or procuring such software requires considerable investment in terms of time, financial resources, and organisational commitment. Consequently, once a department has opted to build a system in-house or to acquire and deploy a commercial solution, it tends to persist in its use, even when the software no longer aligns with the organisation's values⁶⁵.

Developing a system in-house is, however, far from straightforward. Beyond securing a clear and adequate legal basis for the processing of personal data, both for training and for the system's ongoing operation, the public authority must equally possess the requisite technical knowledge and expertise. When these skills are lacking, cooperation with universities or research institutions becomes essential. The MARVEL project illustrates this point well. Regardless of the specific features of the initiative, it is clear that the municipality of Trento did not have the internal capability to develop AI systems

⁶⁴ UNICRI, INTERPOL, *Organizational Roadmap*, cited, 9.

⁶⁵ BRAYNE, *Predict and Surveil*, cited, 27. Similarly, JANSEN, *Data-driven policing*, cited, 71 states that «when police allocate resources to the development of a specific technology and it is supported by internal advocates, it will one way or another become embedded within the organisation».

CONCLUDING REMARKS

on its own. For this reason, it joined a broader project involving several research centres and universities. The case is particularly instructive because it highlights the structural challenges public authorities face when attempting to develop AI tools independently. These challenges become even more pronounced when considering Article 10 of the AI Act, which requires providers to ensure that the datasets used to train, validate and test AI systems are sufficiently representative of the population in relation to which the AI system will be deployed. This implies, in practice, that data should be collected from that specific population, with all the operational, legal, and ethical complexities. The assessment shall take into consideration multiple aspects⁶⁶, such as the indispensability and potential value of the proposed AI use case, including its benefits, risks, and alignment with the agency's strategic and operational goals; evaluate organisational readiness, internal capacity, and the human capital needed to operate and oversee the system; consider the full costs of adoption – including procurement, deployment, and training – against the potential return on investment, such as efficiency gains or process improvements; determine how the AI system can be effectively integrated into the agency's existing technological infrastructure.

Before and during the assessment, the law enforcement authority shall start the engagement with the public and other relevant stakeholders in order to better understand public concerns and help find mutual grounds for acceptability, accountability, transparency, and public buy-in. In Estonia and Sweden, for example, researchers have conducted a study on the public's understanding and perception of predictive policing⁶⁷. The study demonstrates that the integration of data-integration, analysis, and visualisation platforms into police work is perceived by citizens as fundamentally transforming policing practices, particularly by expanding the state's surveillance capacities. Participants expressed a strong awareness of the breadth of data accessible to authorities and voiced scepticism regarding the effectiveness and fairness of predictive policing, which they feared would disproportionately target marginalised groups and reinforce existing social inequalities. The research also highlights concerns that predictive policing may produce counterproductive or even provocative effects, potentially encouraging individuals to engage in criminal behaviour as a form of performativity. Moreover, citizens reported that intrusive or poorly communicated uses of predictive analytics undermine trust in the police, especially in contexts where trust is already fragile. Interestingly, while participants rejected data-driven policing framed around monitoring and control, they were more accepting when such technologies were directed toward enhancing individuals' well-being, for example, by identifying needs related to mental health, social isolation, or unemployment.

⁶⁶ UNICRI, INTERPOL, *Organizational Roadmap*, cited, 21-22.

⁶⁷ A. MASSO, T. KASAPOGLU, A. KAUN, V. GALIS, *Citizens' perspectives on platformisation of police work: a scenario and story-based exploration in Estonia and Sweden*, in *Information, Communication & Society*, 27(13), 2024, 2400-2418, DOI:10.1080/1369118X.2024.2333842.

CONCLUDING REMARKS

The study further reveals various forms of citizen resistance, including complaints, counter-monitoring, and demands for greater transparency, reflecting a broader expectation that predictive technologies should be subject to democratic oversight. In particular, participants in the study – as well as police officers themselves – emphasised the importance of engaging with relevant authorities to shape and influence predictive policing practices. Therefore, establishing a solid foundation for public engagement is a crucial component of responsible AI innovation. Such engagement is essential not only for enhancing transparency and promoting public trust, but also for ensuring that police operations can be carried out effectively and with democratic legitimacy.

As discussed in Chapter 3, the AI Act establishes a risk-based regulatory framework that assigns compliance obligations according to the level of risk an AI system poses, with the overarching aim to mitigate the risks to fundamental rights, reducing them to an acceptable level. While the AI Act explicitly requires providers to implement both a risk-management system⁶⁸ and a quality-management system⁶⁹, deployers should also develop a comprehensive risk-management strategy. Such a strategy ensures that risks are continuously identified, assessed, monitored, and addressed throughout the system's lifecycle, and that the AI system remains aligned with organisational objectives and legal requirements. This strategy should articulate how the agency will identify, assess, respond to, and monitor the risks associated with implementing AI systems, thereby providing a coherent framework for risk governance throughout the AI lifecycle. In this context, risk management involves a coordinated and rigorous process aimed at understanding and addressing potential harms that may arise both from specific AI applications and from the broader organisational use of AI. A robust risk management policy must therefore define the agency's approach to risk assessment, assign roles and responsibilities and allocate the necessary organisational, human, and financial resources. Key elements of this process include assessing risks to individuals and communities, especially through the FRIA discussed in Chapter 4; evaluating additional AI-related risks in accordance with established procedures; developing and implementing informed risk responses; and continuously monitoring risks and mitigation measures at each stage of the AI lifecycle or when relevant circumstances change. As a dynamic and iterative process, risk management requires periodic repetition of assessments and related activities to ensure ongoing oversight, ultimately supporting transparency, accountability, and the minimisation of harm in the deployment of AI systems.

⁶⁸ AI Act, Article 9.

⁶⁹ AI Act, Article 17.

BIBLIOGRAPHY

- Adadi, A., Berrada, M., Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI), in IEEE Access, 6, 2018, 52138-52160.
- Aden, H., Information Sharing, Secrecy and Trust among Law Enforcement and Secret Service Institutions in the European Union, in West European Politics, 41(4), 2018, 981-1002.
- Adinolfi, A., Evoluzione tecnologica e diritti fondamentali: qualche considerazione sulle attuali strategie normative dell'Unione, in I Post di AISDUE, 15, 2023, 321-343.
- Afzal, M., Panagiotopoulos, P., Data in Policing: An Integrative Review, in International Journal of Public Administration, 48(7), 2025, 411-430.
- Agency for Digital Italy (AgID), Guidelines for the Acquisition and Reuse of Software for Public Administrations, February 2020.
- Agency for Digital Italy (AgID), Guidelines for the adoption of Artificial Intelligence in Public Administration, Resolution No. 17/2025, 18 February 2025, forthcoming.
- Agency for Digital Italy (AgID), Guidelines for the development of Artificial Intelligence in Public Administration, 12 March 2026, public consultation.
- Agency for Digital Italy (AgID), Guidelines for the procurement of Artificial Intelligence in Public Administration, 12 March 2026, public consultation.
- Agency for Digital Italy (AgID), Piano Triennale per l'informatica nella Pubblica Amministrazione 2024-2026.
- Agrawal, A., Gans, J., Goldfarb, A., Prediction Machines: The Simple Economics of Artificial Intelligence, Harvard Business Review Press, 2018.
- Aidinlis, S., Barnard-Wills, D., Cochrane, L., Garstka, K., Gurzawska, A., Hughes, J., Between GDPR and Law Enforcement Directive in Security Research: The Use of Personal Data by Law Enforcement Authorities, in European Journal of Law and Technology, 15(3), Article 957, 2024.
- Aiken, C., Classifying AI Systems, Center for Security and Emerging Technology, November 2021.

BIBLIOGRAPHY

- Akhgar, B., Bayerl, P.S., Bailey, K., Dennis, R., Gibson, H., Heyes, S., Lyle, A., Raven, A., Sampson, F., Accountability Principles for Artificial Intelligence (AP4AI) in the Internal Security Domain, Europol, February 2022.
- Allen, D., Karanasios, S., Critical Factors and Patterns in the Innovation Process, in Policing (Oxford): A Journal of Policy and Practice, 5(1), 2011, 87-97.
- Almada, M., Petit, N., The EU AI Act: A Medley of Product Safety and Fundamental Rights?, EUI RSC Working Paper No. 2023/59, 2023.
- Almada, M., Petit, N., The EU AI Act: Between The Rock Of Product Safety And The Hard Place Of Fundamental Rights, in Common Market Law Review, 62, 2025, 85-120.
- Almada, M., Regulation by Design and the Governance of Technological Futures, in European Journal of Risk Regulation, 14(4), 2023, 697-709.
- Álvarez García, V., El proceso de privatización de la calidad y de la seguridad industrial y sus implicaciones desde el punto de vista de la competencia empresarial, in Revista de Administración Pública, 159, 2002, 341-368.
- Álvarez García, V., La confirmación por parte de la jurisprudencia del Tribunal de Justicia de la Unión Europea de la capacidad normativa de los sujetos privados y sus lagunas jurídicas, in Revista General de Derecho Administrativo, 46, 2017.
- Álvarez García, V., The Implementation Of Harmonised Standards And Common Specifications In The Field Of Artificial Intelligence (Articles 40 And 41 AIA), in L. Cotino Hueso, D.-U. Galetta (Eds.), The European Union Artificial Intelligence Act. A Systematic Commentary, CERIDAP, Editoriale Scientifica, 2024, 418-434.
- Ambrose, M.L., Lessons from the Avalanche of Numbers: Big Data in Historical Perspective, in I/S: A Journal of Law and Policy for the Information Society, 11(2), 2015, 201-277.
- Andrejevic, M., Automated Media, Routledge, 2020.
- Angwin, J., Larson, J., Mattu, S., Kirchner, L., Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks, ProPublica, 23 May 2016.
- Antognazza, M.R., Leibniz: An Intellectual Biography, Cambridge University Press, 2009.
- Aradau, C., Blanke, T., Politics of Prediction: Security and the Time/Space of Governmentality in the Age of Big Data, in European Journal of Social Theory, 20(3), 2016, 373-391.

BIBLIOGRAPHY

- Armiento, M.B., *Pubbliche amministrazioni e intelligenza artificiale: Strumenti, principi e garanzie*, Editoriale Scientifica, 2024.
- Article 29 Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679*, 3 October 2017.
- Article 29 Working Party, *Opinion 02/2012 on Facial Recognition in Online and Mobile Services*, March 2012.
- Article 29 Working Party, *Opinion 03/2015 on the Draft Directive on the Protection of Individuals with Regard to the Processing of Personal Data by Competent Authorities for the Purposes of Prevention, Investigation, Detection or Prosecution of Criminal Offences or the Execution of Criminal Penalties, and the Free Movement of Such Data*, 1 December 2015.
- Article 29 Working Party, *Opinion 05/2014 on Anonymisation Techniques*, 10 April 2014.
- Article 29 Working Party, *Opinion 3/2012 on Developments in Biometric Technologies*, 27 April 2012.
- Article 29 Working Party, *Opinion on Some Key Issues of the Law Enforcement Directive (EU 2016/680)*, 29 November 2017.
- Arya, V., Bellamy, R.K.E., Chen, P.-Y. et al., *One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques*, arXiv:1909.03012, 2019.
- Ashworth, A., Zedner, L., *Prevention and Criminalization*, in *New Criminal Law Review: An International and Interdisciplinary Journal*, 15(4), 2012, 542-571.
- Asimov, I., *Robots and Empire*, Harper Voyager, 2023.
- Asimov, I., *Runaround*, in *I, Robot*, Mondadori, 2021.
- Asimov, I., *The Complete Robots*, Harper Voyager, 2018.
- Auby, J.-B., *Il diritto amministrativo di fronte alle sfide digitali*, in *Istituzioni del Federalismo*, 3, 2019, 619-641.
- Avanzini, G., *Decisioni amministrative e algoritmi informatici: Predeterminazione, analisi predittiva e nuove forme di intelleggibilità*, Editoriale Scientifica, 2019.
- Babic, B., Johnson King, Z., *Algorithmic Fairness and Resentment*, in *Philosophical Studies*, 182(1), 2025, 87-119.

BIBLIOGRAPHY

- Babuta, A., Big Data and Policing: An Assessment of Law Enforcement Requirements, Expectations and Priorities, RUSI Occasional Paper, September 2017.
- Babuta, A., Oswald, M., Data Analytics and Algorithms in Policing in England and Wales: Towards a New Policy Framework, RUSI Occasional Paper, February 2020.
- Babuta, A., Oswald, M., Rinik, C., Machine Learning Algorithms and Police Decision-Making: Legal, Ethical and Regulatory Challenges, RUSI Whitehall Report 3-18, 2018.
- Bahceci, B., Redefining the Concept of Penalty in the Case-law of the European Court of Human Rights, in *European Public Law*, 26(4), 2020, 867-888.
- Bakke, E., Predictive Policing: The Argument for Public Transparency, in *NYU Annual Survey of American Law*, 74, 2018, 131-172.
- Baldini, D., De Benedetto, M., The Open Texture of ‘Algorithm’ in Legal Language, in *AI & Society*, 40, 2025, 1643-1654.
- Ballot Jones, L., Thornton, J., De Silva, D., Limitations of Risk-Based Artificial Intelligence Regulation: A Structuration Theory Approach, in *Discover Artificial Intelligence*, 5, Article 14, 2025.
- Balogh, D.A., Near Repeat-Prediction mit PRECOBS bei der Stadtpolizei Zürich, in *Kriminalistik*, 70(5), 2016, 335-341.
- Barkane, I., Questioning the EU Proposal for an Artificial Intelligence Act: The Need for Prohibitions and a Stricter Approach to Biometric Surveillance, in *Information Polity*, 27(2), 2022, 147-162.
- Barnes, G., Focusing Police Resources: Algorithmic Forecasting in Durham, Paper presented to the 9th International Conference on Evidence-Based Policing, Cambridge, July 2016.
- Barnes, G., Hyatt, J.M., Classifying Adult Probationers by Forecasting Future Offending, Final Technical Report, March 2012.
- Barnes, W.H.F., The Nature of Explanation, in *Nature*, 153, 1944, 605.
- Barocas, S., Selbst, A.D., Big Data’s Disparate Impact, in *California Law Review*, 104, 2016, 671-732.
- Barredo Arrieta, A. et al., Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI, in *Information Fusion*, 58, 2020, 82-115.

BIBLIOGRAPHY

- Bartoletti, I., Xenidis, R., Study on the Impact of Artificial Intelligence Systems, Their Potential for Promoting Equality – Including Gender Equality - and the Risks They May Cause in Relation to Non-Discrimination, Council of Europe, 2023.
- Basile, F., Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine, in *Diritto penale e uomo*, 10(2), 2019.
- Bass, S., Policing Space, Policing Race: Social Control Imperatives and Police Discretionary Decisions, in *Social Justice*, 28(1), 2001, 156-176.
- Bassini, M., Liguori, L., Pollicino, O., Sistemi di intelligenza artificiale, responsabilità e accountability. Verso nuovi paradigmi?, in Pizzetti, F. (Ed.), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Giappichelli, 2018.
- Bathae, Y., The Artificial Intelligence Black Box and the Failure of Intent and Causation, in *Harvard Journal of Law & Technology*, 31(2), 2018, 889-938.
- Bayamlioglu, E., Leenes, R., The ‘Rule of Law’ Implications of Data-Driven Decision-Making: A Techno Regulatory Perspective, in *Law, Innovation and Technology*, 10(2), 2018, 295–313.
- Bayamlioglu, E., The Right to Contest Automated Decisions under the General Data Protection Regulation: Beyond the so-Called “Right to Explanation”, in *Regulation & Governance*, 16(4), 2022, 1058-1078.
- Beck, U., *Risk Society: Towards a New Modernity*, Sage Publications, London, 1992.
- Beckley, A., Is Managerialism Alive and Well in the Police Service?, in Albrecht, J.F., Den Heyer, G. (Eds.), *Enhancing Police Service Delivery*, 2021, 341-358.
- Beer, D., Introduction: The Social Power of Algorithms, in Beer, D. (Ed.), *The Social Power of Algorithms*, Routledge, 2018.
- Behn, R.D., *The PerformanceStat Potential: A Leadership Strategy for Producing Results*, Brookings Institution Press, 2014.
- Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J., Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 1997, 711-720.
- Bell, A., Solano-Kamaiko, I., Nov, O., Stoyanovich, J., It’s Just Not That Simple: An Empirical Study of the Accuracy-Explainability Trade-off in Machine Learning for Public Policy, in

BIBLIOGRAPHY

- Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), ACM, New York, 2022, 248-266.
- Bellman, R.E., *An Introduction to Artificial Intelligence: Can Computers Think?*, Boyd & Fraser Publishing Company, 1978.
- Benbouzid, B., To Predict and to Manage: Predictive Policing in the United States, in *Big Data & Society*, 6, 2019, 1-13.
- Bench-Capon, T., Thirty Years of Artificial Intelligence and Law: Editor's Introduction, in *Artificial Intelligence and Law*, 30, 2022, 475-479.
- Bendiek, A., Stuerzer, I., The Brussels Effect, European Regulatory Power and Political Capital, in *Digital Society*, 2, 2023, Article 5.
- Bengio, Y., Courville, A., Vincent, P., Representation Learning: A Review and New Perspectives, arXiv:1206.5538, 2012-2014.
- Benjamin, E., Algorithmic Fairness and Base Rate Tracking, in *Philosophy & Public Affairs*, 50(2), 2022, 239-266.
- Berk, R., Heidari, H., Jabbari, S., Kearns, M., Roth, A., Fairness in Criminal Justice Risk Assessments: The State of the Art, in *Sociological Methods & Research*, 50(1), 2018, 3-44.
- Berk, R.A., Artificial Intelligence, Predictive Policing, and Risk Assessment for Law Enforcement, in *Annual Review of Criminology*, 4, 2021, 209-237.
- Berk, R.A., Sorenson, S.B., Barnes, G., Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions, in *Journal of Empirical Legal Studies*, 13(1), 2016, 94-115.
- Black, J., Baldwin, R., Really Responsive Risk-Based Regulation, in *Law and Policy*, 2010, 32(2), 181-231.
- Black, J., The Role of Risk in Regulatory Processes, in Baldwin, R., Cave, M., Lodge, M. (Eds.), *The Oxford Handbook of Regulation*, Oxford University Press, 2010, 302-348.
- Blandini, A., *Diritto dell'innovazione*, Wolters Kluwer CEDAM, 2022.
- Bledsoe, W.W., Some Results on Multicategory Pattern Recognition, in *Journal of the ACM*, 13(2), 1966, 304-316.

BIBLIOGRAPHY

- Bledsoe, W.W., *The Model Method in Facial Recognition*, Panoramic Research Inc., Palo Alto, CA, Rep. PR1, 1966.
- Bloch-Wehba, H., *Access To Algorithms*, in *Fordham Law Review*, 88(4), 2020, 1265-1314.
- Blount, M., Ebling, M.R., Eklund, J.M. et al., *Real-Time Analysis for Intensive Care: Development and Deployment of the Artemis Analytic System*, in *IEEE Engineering in Medicine and Biology Magazine*, 29(2), 2010, 110-118.
- Bobev, T., *Defining AI in the AI Act: Pin the Tail on the System*, KU Leuven, CiTiP, 2 April 2024.
- Boehm, F., De Hert, P., *Notification, an Important Safeguard Against the Improper Use of Surveillance - Finally Recognized in Case Law and EU Law*, in *European Journal of Law and Technology*, 3(3), 2012.
- Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., Pentland, A., *Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data*, arXiv:1409.2983v1, September 2014.
- Bommasani, R. et al., *On the Opportunities and Risks of Foundation Models*, Center for Research on Foundation Models (CRFM), Stanford University, 2021.
- Boole, G., *The Mathematical Analysis of Logic*, Philosophical Library, New York, 1948.
- Bordt, S., Finck, M., Raidl, E., Von Luxburg, U., *Post-hoc Explanations Fail to Achieve Their Purpose in Adversarial Contexts*, in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, ACM, New York, 2022, 891-905.
- Bostrom, N., *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, 2014.
- Boucher, P.N., *Artificial Intelligence: How Does It Work, Why Does It Matter, and What Can We Do about It?*, Study, Think Tank European Parliament, 2020.
- Bowers, K., Johnson, S.D., Pease, K., *Prospective Hot-Spotting: The Future of Crime Mapping?*, in *British Journal of Criminology*, 44(5), 2004, 641-658.
- Boyd, D., Crawford, K., *Critical Questions for Big Data*, in *Information, Communication and Society*, 15(5), 2012, 662-679.
- Boyle, P., *Securing the Olympic Games: Exemplifications of Global Governance*, in Lenskyj, H.J., Wagg, S. (Eds.), *The Palgrave Handbook of Olympic Studies*, Palgrave Macmillan, 2022, 394-409.

BIBLIOGRAPHY

- Bradford, A., *Digital Empires. The Global Battle to Regulate Technology*, Oxford University Press, 2023.
- Bradford, A., *The Brussels Effect: How the European Union Rules the World*, Oxford University Press, 2020.
- Bradford, A., *The False Choice Between Digital Regulation and Innovation*, in *Northwestern University Law Review*, 119(2), 2024, 377-452.
- Braga, A.A., *Hot Spots Policing: Theoretical Perspectives, Scientific Evidence, and Proper Implementation*, in Teasdale, B., Bradley, M. (Eds.), *Preventing Crime and Violence*, Springer, Cham, 2017, 269-279.
- Braga, A.A., Webster, D.W., White, M.D., Saizow, H., *SMART Approaches to Reducing Gun Violence: Smart Policing Initiative*, 2014.
- Brantingham, P.J., Brantingham, P.L., Andresen, M.A., *The Geometry of Crime and Crime Pattern Theory*, in Wortley, R., Townsley, M. (Eds.), *Environmental Criminology and Crime Analysis*, Routledge, 2017, 98-115.
- Bratton, W., Knobler, P., *Turnaround: How America's Top Cop Reversed the Crime Epidemic*, Random House, 1998.
- Bratton, W., Morgan, J., Malinowski, S., *Fighting Crime in the Information Age: The Promise of Predictive Policing*, LAPD Research Paper, 2009.
- Bratton, W.J., Malinowski, S.W., *Police Performance Management in Practice: Taking COMPSTAT to the Next Level*, in *Policing: A Journal of Policy and Practice*, 2(3), 2008, 259-265.
- Braum, S., Covolo, V., *From Proven Fragmentation to Guaranteed Data Protection within the Virtual Criminal Law Enforcement Area*, in Ligeti, K. (Ed.), *Toward a Prosecutor for the European Union – Volume 1: A Comparative Analysis*, Hart Publishing, 2012.
- Brayne, S., *Big Data Surveillance: The Case of Policing*, in *American Sociological Review*, 82(5), 2017, 977-1008.
- Brayne, S., *Predict and Surveil: Data, Discretion, and the Future of Policing*, Oxford University Press, 2020.
- Breiman, L., *Random Forests*, in *Machine Learning*, 45, 2001, 5-32.

BIBLIOGRAPHY

- Brewczyńska, M., A Critical Reflection on the Material Scope of the Application of the Law Enforcement Directive and its Boundaries with the General Data Protection Regulation, in Kosta, E., Leenes, R., Kamara, I. (Eds.), *Research Handbook on EU Data Protection Law*, Edward Elgar Publishing, 2022, 91-114.
- Brewczyńska, M., Financial Intelligence Units: Reflections on the Applicable Data Protection Legal Framework, in *Computer Law & Security Review*, 43, 2021, Article 105612.
- Brkan, M., Do Algorithms Rule the World? Algorithmic Decision-Making and Data Protection in the Framework of the GDPR and Beyond, in *International Journal of Law and Information Technology*, 27(2), 2019, 91-121.
- Bromley, A.G., Charles Babbage's Analytical Engine, 1838, in *IEEE Annals of the History of Computing*, 20(4), 1998, 29-45.
- Bruveris, M., Gietema, J., Mortazavian, P., Mahadevan, M., Reducing Geographic Performance Differential for Face Recognition, *arXiv:2002.12093v1*, 2020.
- Buchanan, B.G., Shortliffe, E.H., *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, 1994.
- Buijsman, S., Transparency for AI Systems: A Value-Based Approach, in *Ethics and Information Technology*, 26, 2024, Article 34, 1-15.
- Buolamwini, J., Gebru, T., Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, in *Proceedings of Machine Learning Research*, 81, 2018, 1-15.
- Buolamwini, J., Ordóñez, V., Morgenstern, J., Learned-Miller, E., *Facial Recognition Technologies: A Primer*, Algorithmic Justice League, 2020.
- Bureau of Justice Assistance, Police Executive Research Forum, *Compstat: Its Origins, Evolution, and Future in Law Enforcement Agencies*, 2013.
- Burrell, J., How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms, in *Big Data & Society*, 3(1), 2016, 1-12.
- Burt, A., *Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning Models*, The Future of Privacy Forum, 2018.
- Busuioc, M., Curtin, D., Almada, M., Reclaiming Transparency: Contesting the Logics of Secrecy within the AI Act, in *European Law Open*, 2, 2023, 79-98.

BIBLIOGRAPHY

- Bygrave, L.A., Article 22 Automated Individual Decision-Making, Including Profiling, in *The EU General Data Protection Regulation (GDPR)*, Oxford University Press, 2020, 522-542.
- Cabrera, L.L., EU's much-heralded AI Act agreed by EU Parliament - but serious human rights holes in law remain, *Euractiv*, 2024
- Cai, L., Zhu, Y., The Challenges of Data Quality and Data Quality Assessment in the Big Data Era, in *Data Science Journal*, 14, 2015, Article 2, 1-10.
- Calders, T., Verwer, S., Three Naive Bayes Approaches for Discrimination-Free Classification, in *Data Mining and Knowledge Discovery*, 21(2), 2010, 277-292.
- Calders, T., Žliobaitė, I., Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures, in Custers, B., Calders, T., Schermer, B., Zarsky, T. (Eds.), *Discrimination and Privacy in the Information Society*, Springer, 2013, 43-60.
- Campedelli, G.M., *Machine Learning for Criminology and Crime Research: At the Crossroads*, Routledge, 2022.
- Cannataci, J.A., Caruana, M.M., Recommendation R (87) 15 – Twenty-five years down the line, Report, February 2014.
- Cantero Gamito, M., Marsden, C.T., Artificial Intelligence Co-Regulation? The Role of Standards in the EU AI Act, in *International Journal of Law and Information Technology*, 32(1), 2024, Article eaae011.
- Cantero, M., Micklitz, M.W. (Eds.), *The Role of the EU in Transnational Legal Ordering: Standards, Contracts and Codes*, Edward Elgar Publishing, 2020.
- Cao, K., Jain, A.K., Learning Fingerprint Reconstruction: From Minutiae to Image, in *IEEE Transactions on Information Forensics and Security*, 10(1), 2015, 104-117.
- Čapek, K., R.U.R. (Rossum's Universal Robots), Miraggi Edizioni, 2022.
- Caplan, J.M., Kennedy, L.W., *Risk Terrain Modelling Manual: Theoretical Framework and Technical Steps of Spatial Risk Assessment for Crime Analysis*, Rutgers Center on Public Security, Newark, NJ, 2010.
- Carassai, M., Paura e società del rischio. Un'intervista a Ulrich Beck, in *Lo Sguardo*, 21, 2016, 209-218.
- Carbone, A., *Etica digitale nello sviluppo di soluzioni di IA*, Sogei S.p.a., 14 May 2019.

BIBLIOGRAPHY

- Carlini, N., Tramèr, F., Wallace, E. et al., Extracting Training Data from Large Language Models, arXiv, 15 June 2021.
- Caruana, M.M., The Reform of the EU Data Protection Framework in the Context of the Police and Criminal Justice Sector, in *International Review of Law, Computers & Technology*, 33(3), 2017, 249-270.
- Carvalho, D.V., Pereira, E.M., Cardoso, J.S., Machine Learning Interpretability: A Survey on Methods and Metrics, in *Electronics*, 8(8), 2019, Article 832, 1-34.
- Casaburo, D., Marsh, I., Ensuring Fundamental Rights Compliance and Trustworthiness of Law Enforcement AI Systems: The ALIGNER Fundamental Rights Impact Assessment, in *AI and Ethics*, 4, 2024, 1569-1582.
- Casonato, C., Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro, in *Biolaw Journal*, Special Issue, 2, 2019, 711-725.
- Casonato, C., Intelligenza artificiale e diritto costituzionale: prime considerazioni, in *Diritto pubblico comparato ed europeo*, Special Number, 2019, 101-130.
- Casonato, C., L'intelligenza artificiale fra pubblico e privato: una sfida per il costituzionalismo, in *DPCE*, 1, 2025, 5-13.
- Casonato, C., Marchetti, B., Prime osservazioni sulla Proposta di regolamento dell'Unione europea in materia di Intelligenza Artificiale, in *BioLaw Journal*, 3, 2021, 415-437.
- Casonato, C., Per una intelligenza artificiale costituzionalmente orientata, in D'Aloia, A. (Ed.), *Intelligenza artificiale e diritto: Come regolare un mondo nuovo*, FrancoAngeli, 2020, 104-132.
- Casonato, C., Unlocking the Synergy: Artificial Intelligence and (Old and New) Human Rights, in *BioLaw Journal*, 3, 2023, 235-236.
- Cassatella, A., La discrezionalità amministrativa nell'età digitale, in *Diritto amministrativo: scritti per Franco Gaetano Scoca – I*, 2021, 675-708.
- Castellano, P.S., Risk Management Systems as a Specific Obligation for High-Risk Artificial Intelligence Systems in Article 9 of the Regulation, in Cotino Hueso, L., Galetta, D.-U. (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, 2025, 568-594.

BIBLIOGRAPHY

- Castillo Parrilla, J.A., General-Purpose Artificial Intelligence, Foundational Models (And “GPT Chat”) in the Artificial Intelligence Act, in Cotino Hueso, L., Galetta, D.-U. (Eds.), The European Union Artificial Intelligence Act. A Systematic Commentary, CERIDAP, Editoriale Scientifica, 2025, 801-824.
- Cavallo Perin, R., Galetta, D.U. (Eds.), Il diritto dell’amministrazione pubblica digitale, Giappichelli, 2020.
- Cavazos, J. et al., Accuracy Comparison Across Face Recognition Algorithms: Where Are We on Measuring Race Bias?, in IEEE Transactions on Biometrics, Behaviour and Identity Science, 3(1), 2020, 101-111.
- Cavique, L., Implications of Causality in Artificial Intelligence, in Frontiers in Artificial Intelligence, 7, 2024, Article 1339936, 1-4.
- Cavoukian, A., Privacy by Design – The 7 Foundational Principles, May 2010.
- Celeste, E., Digital Constitutionalism: The Role of Internet Bills of Rights, Routledge, 2022.
- CEPEJ, European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment, CEPEJ(2018)14, 3 December 2018.
- Chalapathy, R., Chawla, S., Deep Learning for Anomaly Detection: A Survey, arXiv:1901.03407v2, January 2019.
- Chan, J.B.L., Brereton, D., Legosz, M., Doran, S., E-policing: The Impact of Information Technology on Police Practices, 2001.
- Chan, J.B.L., Governing Police Practice: Limits of the New Accountability, in British Journal of Sociology, 50(2), 1999, 251-270.
- Chan, J.B.L., The Technological Game: How Information Technology Is Transforming Police Practice, in Criminal Justice, 1(2), 2001, 139-159.
- Charniak, E., McDermott, D., Introduction to Artificial Intelligence, Addison-Wesley, 1985.
- Chaveli Donet, E., The Fundamental Rights Impact Assessment By Deployers Of Artificial Intelligence Systems In The Regulation, in Cotino Hueso, L., Galetta, D.-U. (Eds.), The European Union Artificial Intelligence Act. A Systematic Commentary, CERIDAP, Editoriale Scientifica, 2025, 523-566.
- Chen, M. et al., Evaluating Large Language Models Trained on Code, arXiv:2107.03374, 2021.

BIBLIOGRAPHY

- Chen, M., Mao, S., Liu, Y., Big Data: A Survey, in *Mobile Networks and Applications*, 2014, 19, 171-209.
- Chen, R., Chen, X., Ni, B., Ge, Y., Simswap: An Efficient Framework for High Fidelity Face Swapping, *ACM International Conference on Multimedia*, 2020, 2003-2011.
- Chin, J., Lin, L., *Stato di sorveglianza: La via cinese verso una nuova era del controllo sociale*, Bollati Boringhieri, 2022.
- Chouldechova, A., Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments, in *Big Data*, 5(2), 2017, 153-163.
- Christakis, T., Bannelier, K., Castelluccia, C., Le Métayer, D., Mapping the Use of Facial Recognition in Public Spaces in Europe – Part 1-3, Report of the AI-Regulation Chair, MIAI, 2022.
- Citron, D.K., Pasquale, F., The Scored Society: Due Process for Automated Predictions, in 89 *Washington Law Review* 1, 2014, 1-34.
- Civitarese Matteucci, S., Umano troppo umano. Decisioni amministrative automatizzate e principio di legalità, in *Diritto pubblico*, 2019, 1, 5-41.
- Claes, M., The Primacy of EU Law in European and National Law, in Chalmers, D., Arnall, A. (Eds.), *The Oxford Handbook of European Union Law*, Oxford University Press, 2015, 178-211.
- Clarke, A.C., 2001: A Space Odyssey, *Perfection Learning*, 2000.
- Clarke, R., Eck, J.E., *Become a Problem-Solving Crime Analyst*, Routledge, 1st Ed., 2003.
- Clarke, R.A., Information Technology and Dataveillance, in *Communications of ACM*, 31(5), 1988, 498-512.
- Clarke, R.V., Felson, M. (Eds.), *Routine Activity and Rational Choice (Advances in Criminological Theory)*, Routledge, 2018.
- CNIL, Délibération n° 2019-114 du 12 septembre 2019 portant avis sur le projet d'article 9 du projet de loi de finances pour 2020.
- CNIL, Délibération n° 2020-124 du 10 décembre 2020.
- CNIL, Délibération n° 2020-136 du 17 décembre 2020 portant avis sur un projet de décret relatif au recours à la vidéo intelligente pour mesurer le taux de port de masque dans les transports.

BIBLIOGRAPHY

- CNIL, Délibération n° 2022-118 du 8 décembre 2022 portant avis sur un projet de loi portant sur les jeux Olympiques et Paralympiques de 2024.
- CNIL, Facial Recognition: For a Debate Living Up to the Challenges, November 2019.
- Coaffee, J., Evolving Security Motifs, Olympic Spectacle and Urban Planning Legacy, in *Planning Perspectives*, 39(3), 2024, 637-657.
- Cofone, I.N., Algorithmic Discrimination Is an Information Problem, in *Hastings Law Journal*, 70, 2019, 1389-1432.
- Coglianesse, C., A Framework for Governmental Use of Machine Learning, Report to the Admin. Conf. of the U.S., December 2020.
- Coglianesse, C., Administrative Law in the Automated State, in *Daedalus*, 150(3), 2021, 104–120.
- Coglianesse, C., Lai, A., Assessing Automated Administration, in Bullock, J.B. et al. (Eds.), *Oxford Handbook on AI Governance*, 2022.
- Coglianesse, C., Lampmann, E., Contracting for Algorithmic Accountability, in *Administrative Law Review Accord*, 6(2), 2021, 175-206.
- Coglianesse, C., Lehr, D., Regulating by Robot: Administrative Decision Making in the Machine-Learning Era, in *Georgetown Law Journal*, 105, 2017, 1147-1223.
- Cohen, L.E., Felson, M., Social Change and Crime Rate Trends: A Routine Activity Approach, in *American Sociological Review*, 44(4), 1979, 588-608.
- Colombo, C., Eliantonio, M., Harmonized Technical Standards as Part of EU Law, in *Maastricht Journal of European and Comparative Law*, 24(2), 2017, 323-340.
- Colton, K.W., The Impact and Use of Computer Technology by the Police, in *Communications of the ACM*, 22(1), 1979, 10-20.
- Commission Notice – The 'Blue Guide' on the Implementation of EU Product Rules 2022, 2022/C 247/01.
- Committee of Ministers of the Council of Europe, Recommendation CM/Rec(2010)13, The Protection of Individuals with Regard to Automatic Processing of Personal Data in the Context of Profiling, 23 November 2010.

BIBLIOGRAPHY

- Conseil d'État, Avis sur un Projet de Loi relatif aux jeux Olympiques et Paralympiques de 2024, 15 December 2022.
- Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Convention 108), Guidelines on Artificial Intelligence and Data Protection, T-PD(2019)01, 25 January 2019.
- Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Convention 108), Guidelines on Facial Recognition, 2021.
- Contaldo, A., Marotta, L., L'informatizzazione dell'atto amministrativo: cenni sulle problematiche in campo, in *Il Diritto dell'informazione e dell'informatica*, 3, 2002, 571-601.
- Contreras, J.L., *The Cambridge Handbook of Technical Standardization Law. Volume 2*, Cambridge University Press, 2019.
- Corera, M.L., Data And Data Governance And Connections To Data Protection Principles In Article 10 Of The Artificial Intelligence Act, in Cotino Hueso, L., Galetta, D.-U. (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, 595-627.
- Cornish, D.B., Clarke, R.V.G. (Eds.), *The Reasoning Criminal: Rational Choice Perspectives on Offending*, Routledge, 2014.
- Cortese, J.F.N.B., Cozman, F.G., Lucca-Silveira, M.P., Bechara, A.F., Should Explainability Be a Fifth Ethical Principle in AI Ethics?, in *AI and Ethics*, 3, 2023, 123-134.
- Corvalán, J.G., Carro, M.V., Artificial Intelligence Regulation From Outside The European Union, in Cotino Hueso, L., Galetta, D.-U. (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, 77-102.
- Costantino, F., Rischi e opportunità del ricorso delle amministrazioni alle predizioni dei Big Data, in *Diritto pubblico*, 1, 2019, 43-70.
- Cotino Hueso, L., Codes of Conduct, Seals or Certifications for Artificial Intelligence Systems That Are Not High Risk, in Cotino Hueso, L., Galetta, D.-U. (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, 2025, 825-839.
- Cotino Hueso, L., Galetta, D.-U. (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, 2025.

BIBLIOGRAPHY

- Cotino Hueso, L., La primera sentencia del Tribunal de Justicia de la Unión Europea sobre decisiones automatizadas, in *Diario La Ley*, January 2024.
- Cotino Hueso, L., Scope And Delimitation Of High-Risk Systems In The Artificial Intelligence Act, in Cotino Hueso, L., Galetta, D.-U. (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, 2025, 301-326.
- Couchman, H., *Policing by Machines*, Liberty, January 2019.
- Council of Europe, *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, 2024.
- Council of Europe, Recommendation No. R(87)15 of the Committee of Ministers to Member States Regulating the Use of Personal Data in the Police Sector, 17 September 1987.
- Council of Europe, *The European Convention on Human Rights and Policing*, December 2013.
- Council of the European Union, *Conclusions on the Coordinated Plan on Artificial Intelligence*, 6177/19, 11 February 2019.
- Council of the European Union, *European Security Strategy: A Secure Europe in a Better World*, 2009.
- Crawford, K. et al., *AI Now Report*, 2020.
- Crawford, K., *Artificial Intelligence's White Guy Problem*, *The New York Times*, 25 June 2016.
- Crawford, K., *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press, 2021.
- Cukier, K., Mayer-Schönberger, V., *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, 2013.
- Cukier, K.N., Mayer-Schoenberger, V., *The Rise of Big Data: How It's Changing the Way We Think About the World*, in *Foreign Affairs*, 92(3), 2013, 28-40.
- Czarnocki, J., *Will New Definitions of Emotion Recognition and Biometric Data Hamper the Objectives of the Proposed AIA?*, in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, Darmstadt, Germany, 2021, 1-4.
- D'Acquisto, G., *Intelligenza Artificiale: Elementi*, Giappichelli, 2021.

BIBLIOGRAPHY

- D'Albergo, E., Fasciani, T., Giovanelli, G., La governance dell'Intelligenza Artificiale nelle politiche locali, in *Rivista Trimestrale di Scienze dell'Amministrazione*, 4, 2023, 1-26.
- D'Aloia, A., Il diritto verso "il mondo nuovo": Le sfide dell'Intelligenza Artificiale, in D'Aloia, A. (Ed.), *Intelligenza artificiale e diritto: Come regolare un mondo nuovo*, FrancoAngeli, 2020.
- Dantcheva, A., Elia, P., Ross, A., What Else Does Your Biometric Data Reveal? A Survey on Soft Biometrics, in *IEEE Transactions on Information Forensics and Security*, 11(3), 2016, 441-467.
- De Carolis, F., Torchia, L., Marchetti, B. (Eds.), *European Digital Regulation and Its Impact on Member States*, Springer, 2025.
- De Fine Licht, K., De Fine Licht, J., Artificial Intelligence, Transparency, and Public Decision-Making, in *AI & Society*, 35, 2020, 917-926.
- De Gregorio, G., Demkova, S., The Constitutional Right to an Effective Remedy in the Digital Age, in Van Oirsouw, C. et al. (Eds.), *European Yearbook of Constitutional Law*, Springer, 2024, 223-254.
- De Gregorio, G., *Digital Constitutionalism in Europe. Reframing Rights and Powers in the Algorithmic Society*, Cambridge University Press, 2022.
- De Hert, P., Sajfert, J., The fundamental right to personal data protection in criminal investigations and proceedings: framing big data policing through the purpose limitation and data minimisation principles of the Directive (EU) 2016/680, *Brussels Privacy Hub*, 7(31), 2021.
- De Leonardis, O., Book Review: T. Pitch, *La società della prevenzione*, in *Rassegna Italiana di Sociologia*, 1, 2008, 131-133.
- De Silva, D., Alahakoon, D., An Artificial Intelligence Life Cycle: From Conception to Production, in *Patterns*, 3(6), 2022, 100489, 1-13.
- De Vries, A., The Growing Energy Footprint of Artificial Intelligence, in *Joule*, 7(10), 2023, 2191-2194.
- Degadwala, S., Vyas, D., Raval, A.R., Soni, M., Crime Pattern Analysis and Prediction Using Regression Models, 2023 International Conference on Self Sustainable Artificial Intelligence Systems, Erode, India, 2023, 771-776.

BIBLIOGRAPHY

- DEK, Opinion of the Data Ethics Commission. Data Ethics Commission of the Federal Government of Germany, Federal Ministry of the Interior, Building and Community, Federal Ministry of Justice and Consumer Protection, 2019.
- Del Gatto, S., I sistemi proprietari, l'open source e la pubblica amministrazione, in *Giornale di diritto amministrativo*, 5, 2021, 571-577.
- Del Gatto, S., Il riconoscimento facciale: A che punto siamo?, in *Giornale di diritto amministrativo*, 5, 2022, 692-699.
- Delmastro, M., Nicita, A., *Big Data: Come stanno cambiando il nostro mondo*, il Mulino, 2019.
- Den Boer, M., *Juggling the Balance between Preventive Security and Human Rights in Europe*, Security and Human Rights Monitor, DOI: 10.1163/18750230-02602009.
- Di Gennaro, G., Lombardo, E., *Intelligenza artificiale e politiche di sicurezza urbana: verso quali modelli?*, in Uricchio, A.F., Riccio, G., Ruffolo, U. (Eds.), *Intelligenza Artificiale tra etica e diritti*, Cacucci Editore, 2020.
- Di Martino, A., *L'amministrazione per algoritmi ed i pericoli del cambiamento in atto*, in *Diritto dell'Economia*, 3, 2020, 599-634.
- Di Martino, A., *Tecnica e potere nell'amministrazione per algoritmi*, Editoriale Scientifica, 2023.
- Di Paolo, G., Pressacco, L. (Eds.), *Intelligenza Artificiale e Processo Penale. Indagini, Prove, Giudizio*, Università degli Studi di Trento, Quaderni Della Facoltà Di Giurisprudenza, 63, 2022.
- Dieterich, W., Mendoza, C., Brennan, T., *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*, Northpoint Inc, 2016.
- Dimitrova, D., De Hert, P., *The Right of Access Under the Police Directive: Small Steps Forward*, in Medina, M. et al. (Eds.), *Privacy Technologies and Policy: APF 2018. Lecture Notes in Computer Science*, No. 11079, Springer, 111-130.
- Djeffal, C., *Law by Design Obligations: The Future of Regulating Digital Technologies in Europe?*, *Computer Law & Security Review: The International Journal of Technology Law and Practice*, 59, 2025, Article 106232, 1-18.
- Djeffal, C., *The Normative Potential of the European Rule on Automated Decisions: A New Reading for Art. 22 GDPR*, in *Zeitschrift für Ausländisches Öffentliches Recht und Völkerrecht*, 80(4), 2020, 859-893.

BIBLIOGRAPHY

- Doshi-Velez, F. et al., Accountability of AI Under the Law: The Role of Explanation, Berkman Klein Center Working Paper, 2017.
- Doshi-Velez, F., Kim, B., Towards a Rigorous Science of Interpretable Machine Learning, arXiv:1702.08608, 2017.
- Draghi, M., The Future of European Competitiveness, September 2024.
- Drechsler, L., Comparing LED and GDPR Adequacy: One Standard, Two Systems, in *Global Privacy Law Review*, 2020, 1(2), 93-103.
- Dumbrava, C., Artificial Intelligence at EU Borders: Overview of Applications and Key Issues, EPRS, European Parliament, July 2021.
- Dushi, D., The Use of Facial Recognition Technology in EU Law Enforcement: Fundamental Rights Implications, *Global Campus of Human Rights*, 2020.
- Ebers, M., Hoch, V.R.S., Rosenkranz, F., Ruschemeier, H., Steinrötter, B., The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS), in *J*, 4, 2021, 589-603.
- Ebers, M., Standardizing AI – The Case of the European Commission's Proposal for an Artificial Intelligence Act, in L.A. DiMatteo, C. Poncibò, M. Cannarsa, (Eds.), *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics*, Cambridge University Press, 2022, 321-344.
- Eck, J., Spelman, W., Problem Solving: Problem-Oriented Policing in Newport News, *Police Executive Research Forum*, 1987.
- Edwards, L., Veale, M., Enslaving the Algorithm: From a “Right to an Explanation” to a “Right to Better Decisions”?, in *IEEE Security & Privacy*, 16(3), 2018, 46-54.
- Edwards, L., Veale, M., Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For, in *Duke Law & Technology Review*, 16, 2017, 18-84.
- Egbert, S., Leese, M., *Criminal Futures: Predictive Policing and Everyday Police Work*, Routledge, 2021.
- Egbert, S., Predictive Policing and the Platformization of Police Work, in *Surveillance & Society*, 17(1/2), 2019, 83-88.

BIBLIOGRAPHY

- Eisenhut, D., Sovereignty, National Security and International Treaty Law. The Standard of Review of International Courts and Tribunals with regard to “Security Exceptions”, in *Archiv Des Völkerrechts*, 48(4), 2010, 431–466.
- Eliantonio, M., Cauffman, C., The Legitimacy of Standardisation as a Regulatory Technique - A Cross Disciplinary and Multi-level Analysis: An Introduction, in M. Eliantonio, C. Cauffman (Eds.), *The Legitimacy of Standardisation as a Regulatory Technique - A Cross Disciplinary and Multi-level Analysis*, Edward Elgar Publishing, 2020, 1-18.
- Ellul, J., Should We Regulate Artificial Intelligence or Some Uses of Software?, in *Discover Artificial Intelligence*, 2, 2022, Article 5.
- Engstrom, D.F., Ho, D.E., Sharkey, C.M., Cuéllar, M.-F., Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies, Report Submitted to the Administrative Conference of the United States, February 2020.
- Ericson, R.V., Haggerty, K.D., *Policing the Risk Society*, University of Toronto Press, 1997.
- Ericson, R.V., Shearing, C.D., The Scientification of Police Work, in G. Böhme, N. Stehr (Eds.), *The Knowledge Society: The Growing Impact of Scientific Knowledge on Social Relations*, Reidel, 1986, 129-159.
- Escajedo San-Epifanio, L., Biometric Recognition In The Artificial Intelligence Act: Exemptions, Prohibitions And High-Risk Specialties, in L. Cotino Hueso, D.-U. Galetta (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, 2025, 183-228.
- Esposito, L., Trasparenza e Accountability nelle Amministrazioni Pubbliche: analisi del significato valoriale dei concetti di Trasparenza e Accountability in un'ottica di Open Government, in www.ildirittoamministrativo.it, 4 February 2021.
- Eterno, J.A., Silverman, E.B., *The Crime Numbers Game: Management by Manipulation*, CRC Press, 2012.
- ETSI, Securing Artificial Intelligence (SAI). Traceability of AI Models, ETSI TR 104 032 V1.1.1, February 2024.
- European Commission against Racism and Intolerance (ECRI), General Policy Recommendation N°11 on Combating Racism and Racial Discrimination in Policing, 4 October 2007.

BIBLIOGRAPHY

- European Commission, Commission Implementing Decision on a Standardisation Request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in Support of Union Policy on Artificial Intelligence, C(2023)3215, 22 May 2023.
- European Commission, Commission Implementing Decision on a Standardisation Request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in Support of Union Policy on Artificial Intelligence, C(2023)3215, 22 May 2023.
- European Commission, Commission Work Programme 2025, COM(2025) 45 final, 11 February 2025.
- European Commission, Communication: A European Strategy for Data, COM(2020) 66 final, 19 February 2020.
- European Commission, Communication: Artificial Intelligence for Europe, COM(2018) 237 final, 25 April 2018.
- European Commission, Communication: Building Trust in Human-Centric Artificial Intelligence, COM(2019) 168 final, 8 April 2019.
- European Commission, Communication: Commission Guidelines on Prohibited Artificial Intelligence Practices Established by Regulation (EU) 2024/1689 (AI Act), C(2025) 5052 final, 29 July 2025.
- European Commission, Communication: Coordinated Plan on Artificial Intelligence, COM(2018) 795 final, 7 December 2018.
- European Commission, Communication: First Report on Application and Functioning of the Data Protection Law Enforcement Directive (EU) 2016/680 ('LED'), COM(2022) 364 final, 25 July 2022.
- European Commission, Communication: Fostering A European Approach To Artificial Intelligence, COM(2021) 205 final, 21 April 2021.
- European Commission, Digital Compass 2030, COM(2021) 118 final, 9 March 2021.
- European Commission, Draft Standardisation Request to the European Standardisation Organisations in Support of Safe and Trustworthy Artificial Intelligence, 5 December 2022.

BIBLIOGRAPHY

- European Commission, Guidelines on Prohibited Artificial Intelligence Practices Established by Regulation (EU) 2024/1689 (AI Act), COM(2025) 884 final, 4 February 2025.
- European Commission, Guidelines on the Definition of an Artificial Intelligence System Established by Regulation (EU) 2024/1689 (AI Act), COM(2025) 924 final, 6 February 2025.
- European Commission, Impact Assessment of the Regulation on Artificial Intelligence, SWD(2021) 84 final, 2021.
- European Commission, Strategy on the Software Open Source 2020-2023, COM(2020) 7149 final, 21 October 2020.
- European Commission, White Paper on Artificial Intelligence - A European Approach to Excellence and Trust, COM(2020) 65 final, 19 February 2020.
- European Data Protection Board – European Data Protection Supervisor, Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), 18 June 2021.
- European Data Protection Board, Contribution of the EDPB to the European Commission's Evaluation of the Data Protection Law Enforcement Directive (LED) under Article 62, 14 December 2021.
- European Data Protection Board, Guidelines 05/2022 on the Use of Facial Recognition Technology in the Area of Law Enforcement, 26 April 2023.
- European Data Protection Board, Guidelines 3/2019 on Processing of Personal Data through Video Devices, 2020.
- European Data Protection Board, Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679, 2018.
- European Data Protection Board, Guidelines on Data Protection by Design and by Default, 20 October 2020.
- European Data Protection Board, Opinion 11/2024 on the Use of Facial Recognition to Streamline Airport Passengers' Flow (Compatibility with Articles 5(1)(e) and (f), 25 and 32 GDPR), 24 May 2024.
- European Data Protection Board, Opinion 28/2024 on Certain Data Protection Aspects Related to the Processing of Personal Data in the Context of AI Models, 17 December 2024.

BIBLIOGRAPHY

- European Data Protection Supervisor, Guidance for Risk Management of Artificial Intelligence Systems, 11 November 2025.
- European Data Protection Supervisor, Opinion 6/2015: A Further Step towards Comprehensive EU Data Protection, 28 October 2015.
- European Data Protection Supervisor, TechDispatch #2/2023 on Explainable Artificial Intelligence, 16 November 2023.
- European Law Institute, The Concept of 'AI System' under the New AI Act: Arguing for a Three-Factor Approach, December 2024.
- European Parliament, Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).
- European Parliament, Resolution of 20 October 2020 on Intellectual Property Rights for the Development of Artificial Intelligence Technologies (2020/2015(INI)).
- European Parliament, Resolution of 20 October 2020 with Recommendations to the Commission on a Civil Liability Regime for Artificial Intelligence (2020/2014(INL)).
- European Parliament, Resolution of 20 October 2020 with Recommendations to the Commission on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies (2020/2012(INL)).
- European Parliament, Resolution of 6 October 2021 on Artificial Intelligence in Criminal Law and Its Use by Law Enforcement Authorities in Criminal Matters (2020/2016(INI)).
- Europol, Accountability Principles for Artificial Intelligence (AP4AI) in the Internal Security Domain, February 2022.
- Europol, AI and Policing: The Benefits and Challenges of Artificial Intelligence for Law Enforcement, Report from the Europol Innovation Lab, Publications Office of the European Union, 2024.
- Ezzeddine, Y., Artificial Intelligence in Law Enforcement Surveillance: Citizen Perspectives, Resistance and Counterstrategies, Doctoral Thesis, Sheffield Hallam University, 2024.
- Falletta, P., Marsano, A., Intelligenza artificiale e protezione dei dati personali: il rapporto tra Regolamento europeo sull'intelligenza artificiale e GDPR, in *Rivista italiana di informatica e diritto*, 6(1), 2024, 120-137.

BIBLIOGRAPHY

- Falletti, E., Automated Decisions and Article 22 GDPR of the European Union: An Analysis of the Right to an “Explanation”, in *Osservatorio dei diritti fondamentali in Europa*, 31, 2019, 1-15.
- Fantigrossi, U., *Automazione e pubblica amministrazione: Profili giuridici*, il Mulino, 1993.
- Feigenbaum, E., Shrobe, H., The Japanese National Fifth Generation Project: Introduction, Survey, and Evaluation, in *Future Generation Computer Systems*, 9(2), 1993, 105-117.
- Feldstein, S., *The Global Expansion of AI Surveillance*, Working Paper, Carnegie Endowment for International Peace, September 2019.
- Ferguson, A.G., Illuminating Black Data Policing, in *Ohio State Journal of Criminal Law*, 15, 2018, 503-525.
- Ferguson, A.G., Policing Predictive Policing, in *Washington University Law Review*, 94(5), 2017, 1109-1189.
- Ferguson, A.G., Predictive Policing and Reasonable Suspicion, in *Emory Law Journal*, 62(2), 2012, 259-325.
- Ferguson, A.G., *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, NYU Press, 2017.
- Fernández Llorca, D. et al., An Interdisciplinary Account of the Terminological Choices by EU Policymakers ahead of the Final Agreement on the AI Act: AI System, General Purpose AI System, Foundation Model, and Generative AI, in *Artificial Intelligence and Law*, 33(4), 2024, 875–888.
- Fernandez, J.V., Artificial Intelligence in Government: Risks and Challenges of Algorithmic Governance in the Administrative State, in *Indiana Journal of Global Legal Studies*, 30(1), 2023, 65-120.
- Ferrarese, M.R., Digitalizzazione e diritto. Alcune domande e questioni, in M. Ramajoli (Ed.), *Una giustizia amministrativa digitale?*, il Mulino, 2023, 23-37.
- Fidler, M., Hurwitz, J., An Overview of Facial Recognition Technology Regulation in the United States, in R. Matulionyte, M. Zalnieriute (Eds.), *The Cambridge Handbook of Facial Recognition in the Modern State*, Cambridge Law Handbooks, Cambridge University Press, 2024, 214-227.

BIBLIOGRAPHY

- Finocchiaro, G., Anonimato, in *Digesto Discipline Privatistiche – Sezione Civile, Aggiornamento V*, 2010.
- Finocchiaro, G., *Intelligenza Artificiale e protezione dei dati personali*, in *Giurisprudenza italiana*, 2019, 1670–1677.
- Finocchiaro, G., *Intelligenza Artificiale, quali regole?*, il Mulino, 2024.
- Floridi, L., Cows, J., *A Unified Framework of Five Principles for AI in Society*, in L. Floridi (Ed.), *Ethics, Governance, and Policies in Artificial Intelligence*, Springer, 2021, 5-17.
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., Vayena, E., *AI4People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, in *Minds and Machines*, 28(4), 2018, 689-707.
- Floridi, L., *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Raffaello Cortina Editore, 2022.
- Floridi, L., *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Raffaello Cortina Editore, 2017.
- Floridi, L., *Pensare l'infosfera. La filosofia come design concettuale*, Raffaello Cortina Editore, 2020.
- Floridi, L., *Soft Ethics and the Governance of the Digital*, in *Philosophy & Technology*, 31, 2018, 1-8.
- Floridi, L., *Soft Ethics: Its Application to the General Data Protection Regulation and Its Dual Advantage*, in *Philosophy & Technology*, 31, 2018, 163-167.
- Foucault, M., *Sécurité, territoire, population: Cours au Collège de France 1977-1978*, Hautes Etudes, 2004.
- Foucault, M., *Surveiller et punir: Naissance de la prison*, Gallimard, 1975.
- Fraser, H., Bello y Villarino, J.-M., *Acceptable Risks in Europe's Proposed AI Act: Reasonableness and Other Principles for Deciding How Much Risk Management Is Enough*, in *European Journal of Risk Regulation*, 14(2-3), 2023, 352-371.
- Friedman, B., Nissenbaum, H., *Bias in Computer Systems*, in *ACM Transactions on Information Systems*, 14(3), 1996, 330-347.

BIBLIOGRAPHY

- Fundamental Rights Agency (FRA), Bias in Algorithms - Artificial Intelligence and Discrimination, Publications Office of the European Union, December 2022.
- Fundamental Rights Agency (FRA), Crime, Safety and Victims' Rights – Fundamental Rights Survey, Report, February 2021.
- Fundamental Rights Agency (FRA), Data Quality and Artificial Intelligence – Mitigating Bias and Error to Protect Fundamental Rights, Publications Office of the European Union, 2019.
- Fundamental Rights Agency (FRA), Facial Recognition Technology: Fundamental Rights Considerations in the Context of Law Enforcement, Publications Office of the European Union, 2019.
- Fundamental Rights Agency (FRA), Getting the Future Right - Artificial Intelligence and Fundamental Rights, Publications Office of the European Union, 2020.
- Fundamental Rights Agency (FRA), Guidance on Preventing Unlawful Profiling Now and in the Future, Publications Office of the European Union, 2019.
- Fundamental Rights Agency (FRA), Handbook on European Non-Discrimination Law – 2018 Edition, 21 March 2018.
- Fundamental Rights Agency (FRA), Second European Union Minorities and Discrimination Survey, Publications Office of the European Union, 2018.
- Fundamental Rights Agency (FRA), Surveillance by Intelligence Services: Fundamental Rights Safeguards and Remedies in the EU - 2023 Update, May 2023.
- Fundamental Rights Agency (FRA), Your Rights Matter: Police Stops – Fundamental Rights Survey, Publications Office of the European Union, May 2021.
- Fussey, P., Murray, D., Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology, Human Rights Centre, University of Essex, 2019.
- Fussey, P., Sandhu, A., Surveillance Arbitration in the Era of Digital Policing, in Theoretical Criminology, 26(1), 2021, 3-22.
- Future of Life Institute, Asilomar AI Principles, Beneficial AI 2017 Conference, 2017.
- Future of Privacy Forum, The Privacy Expert's Guide to Artificial Intelligence and Machine Learning, 2018.

BIBLIOGRAPHY

G20 Ministerial Statement on Trade and Digital Economy, 9 June 2019.

Galbally, J., Ferrara, P., Haraksim, R., Psyllos, A., Beslay, L., Study on Face Identification Technology for Its Implementation in the Schengen Information System, Joint Research Centre JRC-34751, July 2019.

Galetta, A., The Changing Nature of the Presumption of Innocence in Today's Surveillance Societies: Rewrite Human Rights or Regulate the Use of Surveillance Technologies?, in *European Journal of Law and Technology*, 4(2), 2013.

Galetta, D.U., Corvalán, J.G., Intelligenza artificiale per un'amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto, in *Federalismi.it*, 3, 2019, 1-23.

Galis, V., Karlsson, B., A World of Palantir – Ontological Politics in the Danish Police's POL-INTEL, in *Information, Communication & Society*, 27(13), 2024, 2438-2456.

Gantz, J., Reinsel, D., Extracting Value from Chaos, IDC Report, 2011.

Garante per la Protezione dei Dati Personali (Pasquale Stanzone, Pres.), Segnalazione al Parlamento e al Governo sull'Autorità per l'I.A., 25 March 2024.

Garante per la Protezione dei Dati Personali, Decision No. 127/2021, 25 March 2021.

Garante per la Protezione dei Dati Personali, Decision No. 440/2018, 26 July 2018.

Garante per la Protezione dei Dati Personali, Decision No. 54/2020, 26 February 2020.

Garante per la Protezione dei Dati Personali, Decision of 11 January 2024, No. 5.

Garante per la Protezione dei Dati Personali, Ordinanza ingiunzione nei confronti di Clearview AI, 10 February 2022.

Garante per la Protezione dei Dati Personali, Video Surveillance Decision, 8 April 2010.

Garapon, A., Lassègue, J., Justice Digitale. Révolution graphique et rupture anthropologique, Presses Universitaires de France, Paris, 2018.

Garland, D., The Culture of Control: Crime and Social Order in Contemporary Society, University of Chicago Press, 2001.

Garnett, K., Parsons, D.J., Multi-Case Review of the Application of the Precautionary Principle in European Union Law and Case Law, in *Risk Analysis*, 37(3), 2017, 502-516.

BIBLIOGRAPHY

- Garvie, C., Garbage In, Garbage Out: Face Recognition on Flawed Data, Georgetown Law, Center on Privacy and Technology, May 2019.
- Gates, K., Policing as Digital Platform, in *Surveillance & Society*, 17(1/2), 2019, 63-68.
- Gates, K.A., *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance*, NYU Press, 2011.
- Gelernter, H., Realization of a Geometry-Theorem Proving Machine, in *Proceedings of the International Conference on Information Processing*, 1959, 273-282.
- Gelman, A., Bayesian statistics: What's it all about?, Blog post, 13 December 2016.
- Gerards, J., Xenidis, R., *Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non-Discrimination Law*, Publications Office of the European Union, 2021.
- Gerstner, D., Predictive Policing in the Context of Residential Burglary: An Empirical Illustration on the Basis of a Pilot Project in Baden-Württemberg, Germany, in *European Journal for Security Research*, 3, 2018, 115-138.
- Giannini, M.S., *Il potere discrezionale della pubblica amministrazione: Concetto e problemi*, Giuffrè, 1939.
- Gikay, A.A., Regulating Use by Law Enforcement Authorities of Live Facial Recognition Technology in Public Spaces: An Incremental Approach, in *Cambridge Law Journal*, 82(3), 2023, 414,-449.
- Glauner, P., An Assessment of the AI Regulation Proposed by the European Commission, in S. Ehsani, P. Glauner, P. Plugmann, F.M. Thieringer (Eds.), *The Future Circle of Healthcare: AI, 3D Printing, Longevity, Ethics, and Uncertainty Mitigation*, Springer, 2022, 119-127.
- Goddard, K. et al., Automation Bias: A Systemic Overview of Frequency, Effect Mediators, and Mitigators, in *Journal of the American Medical Informatics Association*, 19(1), 2012, 121-127.
- Goldstein, H., Improving Policing: A Problem-Oriented Approach, in *Crime and Delinquency*, 25(2), 1979, 236-258.
- Goldstein, H., *Problem-Oriented Policing*, McGraw-Hill, 1990.
- Golia, A.J., Critique of Digital Constitutionalism: Deconstruction and Reconstruction from a Societal Perspective, in *Global Constitutionalism*, 13(3), 2024, 488-518.

BIBLIOGRAPHY

- González Fuster, G., Artificial Intelligence and Law Enforcement: Impact on Fundamental Rights, Study commissioned by the European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs, July 2020.
- Goodfellow, I., Bengio, Y., Courville, A., Deep Learning, MIT Press, Boston, 2016.
- Goodman, B., Flaxman, S., European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation", in AI Magazine, 38(3), 2017, 50-57.
- Gornet, M., Maxwell, W., The European Approach to Regulating AI through Technical Standards, in Internet Policy Review, 13(3), 2024.
- Gornet, M., Maxwell, W., The European Approach to Regulating AI through Technical Standards, in Internet Policy Review, 13(3), 2024.
- Green, B., Hu, L., The Myth in the Methodology: Towards a Recontextualization of Fairness in Machine Learning, Presented at the Machine Learning: The Debates Workshop at the 35th International Conference on Machine Learning, Stockholm, Sweden, 2018.
- Gregory, R.L., Intelligence, in The Oxford Companion to the Mind, Oxford University Press, 1987, 375-379.
- Grijpink, J., Privacy Law: Biometrics and Privacy, in Computer Law & Security Report, 17(3), 2001, 154-160.
- Grochowski, J., Jabłowska, J., Lagioia, F., Sartor, G., Algorithmic Transparency and Explainability for EU Consumer Protection, in Critical Analysis of Law, 8(1), 2021, 43-63.
- Grother, P., Ngan, M., Hanaoka, K. (NIST), Face Recognition Vendor Test, Part 3: Demographic Effects, Report, NISTIR 8280, December 2019.
- Gruetzemacher, R., Whittlestone, J., The Transformative Potential of Artificial Intelligence, in Futures, 135, 2022, 1-11.
- Guastafarro, B., Beyond the Exceptionalism of Constitutional Conflicts: The Ordinary Functions of the Identity Clause, in Yearbook of European Law, 31(1), 2012, 263-318.
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D., A Survey Of Methods For Explaining Black Box Models, in ACM Computing Surveys, 51(5), 2018, Article No. 93, 1-42.

BIBLIOGRAPHY

- Gültekin-Várkonyi, G., Predictive Policing and Bias in a Nutshell: Technical and Practical Aspects of Personal Data Processing for Law Enforcement Purposes, in K. Karsai, A. Sözüer, L. Wörner (Eds.), *Digital Criminal Justice: A Studybook Selected Topics for Learners and Researchers*, 2022, 39-59.
- Gundhus, H.O.I., Skjevraak, P.E., Wathne, C.T., We Will Always Be Better Than a Spreadsheet: Intelligence Logic and Crime Prevention in Practice, in *European Journal of Policing Studies*, 6(1), 2023, 27-49.
- Gundhus, H.O.I., Wathne, C.T., Resistance to Platformization: Palantir in the Norwegian Police, in *Information, Communication & Society*, 27(13), 2024, 2457-2475.
- Gunning, D., Aha, D.W., DARPA's Explainable Artificial Intelligence Program, in *AI Magazine*, 40(2), 2019, 44-58.
- Guo, Y., Zhang, L., Hu, Y., He, X., Gao, J., Ms-Celeb-1m: A Dataset and Benchmark for Large-Scale Face Recognition, *European Conference on Computer Vision*, 2016, 87-102.
- Gutiérrez David, M.E., Transparency and Provision of Information to Deployers in Article 13, in L. Cotino Hueso, D.-U. Galetta (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, Napoli, 2024, 668-714.
- Gyevnar, B., Ferguson, N., Schafer, B., Bridging the Transparency Gap: What Can Explainable AI Learn from the AI Act?, in *Frontiers in Artificial Intelligence and Applications*, 372, 2023, 964-965.
- Hacker, P., Comments on the Final Trilogue Version of the AI Act, *European New School of Digital Studies*, 23 January 2024.
- Hacker, P., Passoth, J.H., Varieties of AI Explanations Under the Law. From the GDPR to the AIA, and Beyond, in A. Holzinger, R. Goebel, R. Fong, T. Moon, K.R. Müller, W. Samek (Eds.), *xxAI - Beyond Explainable AI. xxAI 2020, Lecture Notes in Computer Science*, Springer, Cham, Volume 13200, 2022, 343-381.
- Hacker, P., Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law, in *Common Market Law Review*, 55(4), 2018, 1143-1185.
- Hakeem, M., The Validity of the Burgess Method of Parole Prediction, in *American Journal of Sociology*, 53(5), 1948, 376-386.

BIBLIOGRAPHY

- Hamburg DPA, Datenschutzrechtliche Prüfung des Einsatzes einer Gesichtserkennungssoftware zur Aufklärung von Straftaten im Zusammenhang mit dem G20-Gipfel durch die Polizei Hamburg, 31 August 2018.
- Hamilton, M., Predictive Policing through Risk Assessment, in J.L.M. McDaniel, K.G. Pease (Eds.), *Predictive Policing and Artificial Intelligence*, Routledge, 2021, 57-77.
- Han, J., Pei, J., Kamber, M., *Data mining: concepts and techniques*, Elsevier, 2011.
- Hansen, M., Article 27 Data Protection Impact Assessment, in E. Kosta, F. Boehm (Eds.), *The EU Law Enforcement Directive (LED): A Commentary*, Oxford University Press, 2024, 417-440.
- Hanson, B., Stall, S., Cutcher Gershenfeld, J., Vrouwenvelder, K., Wirz, C., Rao, Y., Peng, G., *Garbage In, Garbage Out: Mitigating Risks and Maximizing Benefits of AI in Research*, in *Nature*, 623(2), 2023, 28-31.
- Harcourt, B., The Shaping of Chance: Actuarial Models and Criminal Profiling at the Turn of the Twenty-First Century, in *University of Chicago Law Review*, 70(1), 2003, 105-128.
- Harcourt, B.E., *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*, University of Chicago Press, 2008.
- Harmon, R., Why Do We (Still) Lack Data on Policing?, in *Marquette Law Review*, 96(4), 2013, 1119-1146.
- Hassan, B., Izquierdo, E., Piatrik, T., *Soft Biometrics: A Survey, Benchmark Analysis, Open Challenges and Recommendations*, in *Multimedia Tools and Applications*, 83, 2024, 15151-15194.
- Hastie, T., Tibshirani, R., Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2009.
- Haugeland, J., *Artificial Intelligence: The Very Idea*, MIT Press, 1985.
- Heaven, W.D., Predictive policing algorithms are racist. They need to be dismantled, *MIT Technology Review*, 2020.
- Hedden, B., On Statistical Criteria of Algorithmic Fairness, in *Philosophy & Public Affairs*, 49(2), 2021, 214-231.
- High-Level Expert Group on Artificial Intelligence, *A Definition of AI: Main Capabilities and Disciplines*, December 2018.

BIBLIOGRAPHY

- High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, April 2019.
- High-Level Expert Group on Artificial Intelligence, Policy and Investment Recommendations for Trustworthy AI, June 2019.
- Hildebrandt, M., Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology, Edward Elgar, 2015.
- Hodgkinson, I.R., Jennings, N., Jackson, T., Everyone Must Understand the Environmental Costs of AI, OECD.AI, October 2024.
- Hopfield, J.J., Neural Networks and Physical Systems with Emergent Collective Computational Abilities, in Proceedings of the National Academy of Sciences of the USA, 79(8), 1982, 2554-2558.
- House of Lords, Justice and Home Affairs Committee, Technology Rules? The Advent of New Technologies in the Justice System, HL Paper 180, London, 30 March 2022.
- Human Rights Council, Surveillance and Human Rights: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, United Nations, 2019.
- Huq, A.Z., A Right to a Human Decision, in Virginia Law Review, 106(3), 2020, 611-688.
- Icove, D.J., Automated Crime Profiling, in FBI Law Enforcement Bulletin, 55(12), 1986, 27-30.
- IEEE Standard for Transparency of Autonomous Systems, IEEE 7001-2021.
- IEEE, Ethically Aligned Design, First Edition, IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2019.
- Information Commissioner's Office, Alan Turing Institute, Explaining Decisions Made with AI, 2022.
- Information Commissioner's Office, ICO Finds Metropolitan Police Service's Gangs Matrix Breached Data Protection Laws, 2018.
- Information Commissioner's Office, Information Commissioner's Opinion: The Use of Live Facial Recognition Technology in Public Places, June 2021.

BIBLIOGRAPHY

- Inglese, M., Il regolamento sull'intelligenza artificiale come atto per il completamento e il buon funzionamento del mercato interno?, in Quaderni AISDUE - Rivista quadrimestrale, Fascicolo speciale, 2(2), 2024, 1-25.
- Introna, L., Nissenbaum, H., Facial Recognition Technology: A Survey of Policy and Implementation Issues, Lancaster University Management School Working Paper 2010/030, 2010.
- IEC 31010:2019, Risk management — Risk assessment techniques.
- ISO 31000:2018, Risk management — Guidelines
- ISO/IEC 12792:2025, Information technology — Artificial intelligence (AI) — Transparency taxonomy of AI systems.
- ISO/IEC 19792:2009, Information technology — Security techniques — Security evaluation of biometrics
- ISO/IEC 19794-5:2011, Information technology — Biometric data interchange formats — Part 5: Face image data
- ISO/IEC 22989:2022, Information technology — Artificial intelligence — Artificial intelligence concepts and terminology
- ISO/IEC 2382-37:2022, Information technology — Vocabulary — Part 37: Biometrics .
- ISO/IEC 23894:2023, Information technology — Artificial intelligence — Guidance on risk management
- ISO/IEC 27005:2022, Information security, cybersecurity and privacy protection — Guidance on managing information security risks.
- ISO/IEC 38507:2022, Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations
- ISO/IEC 39794-5:2019, Information technology — Extensible biometric data interchange formats — Part 5: Face image data
- ISO/IEC 42001:2023, Information technology — Artificial intelligence — Management system
- ISO/IEC 5055:2021, Information technology — Software measurement — Software quality measurement — Automated source code quality measures
- ISO/IEC Guide 51:2014, Safety aspects — Guidelines for their inclusion in standards

BIBLIOGRAPHY

ISO/IEC/IEEE 24765:2017, Systems and software engineering — Vocabulary

Jansen, F., Data Driven Policing in the Context of Europe, Working Paper, Datajustice, 2018.

Jansen, F., Data-Driven Policing: Negotiating the Legitimacy of the Police, Doctoral Thesis, Department of Journalism, Media and Culture Cardiff University, January 2022.

Jasserand, A.C., Avoiding Terminological Confusion between the Notions of 'Biometrics' and 'Biometric Data': An Investigation into the Meanings of the Terms from a European Data Protection and a Scientific Perspective, in International Data Privacy Law, 6(1), 2016, 63-76.

Jasserand-Breeman, C., Reprocessing of Biometric Data for Law Enforcement Purposes: Individuals' Safeguards Caught at the Interface between the GDPR and the 'Police' Directive, Doctoral Thesis, University of Groningen, March 2019.

Jobin, A., Ienca, M., Vayena, E., The Global Landscape of AI Ethics Guidelines, in Nature Machine Intelligence, 1, 2019, 389-399.

Joint Research Centre (JRC), Artificial Intelligence: A European Perspective, Publications Office of the European Union, 2018.

Joint Research Centre (JRC), Nativi, S., De Nigris, S., AI Standardisation Landscape, Publications Office of the European Union, 2021.

Joint Research Centre (JRC), Tangi, L. et al., AI Watch European Landscape on the Use of Artificial Intelligence by the Public Sector, JRC Science For Policy Report, European Union, 2022.

Kachur, A., Osin, E., Davydov, D. et al., Assessing the Big Five Personality Traits Using Real-Life Static Facial Images, in Scientific Reports, 10(1), 2020, Article 8487, 1-11.

Kaminski, M.E., Malgieri, G., The Right to Explanation in the AI Act, U of Colorado Law Legal Studies Research Paper No. 25-9, 2025.

Kaminski, M.E., The Right to Explanation, Explained, in Berkeley Technology Law Journal, 34, 2019, 189-218.

Kaminski, M.E., Urban, J.M., The Right to Contest AI, in Columbia Law Review, 121, 2021, 1957-2048.

Kaminsky, M., Regulating the Risks of AI, in Boston University Law Review, 103, 2023, 1347-1373.

BIBLIOGRAPHY

- Kamiran, F., Calders, T., Pechenizkiy, M., Techniques for Discrimination-Free Predictive Models, in B. Custers, T. Calders, B. Schermer, T. Zarsky (Eds.), *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases*, Springer, 2013, 223-240.
- Kanade, T., *Picture Processing System by Computer Complex and Recognition of Human Faces*, Doctoral Thesis, Kyoto University, 1974.
- Kanwal, R., Walby, K., *Tracking the Surveillance and Information Practices of Data Brokers: A Report*, 4 July 2024.
- Kearns, I., Muir, R., *Data-driven Policing and Public Value*, Police Foundation, March 2019.
- Kellerbauer, M., Klamert, M., Tomkin, J., Treaty on the Functioning of the European Union, in M. Kellerbauer, M. Klamert, J. Tomkin (Eds.), *The EU Treaties and Charter of Fundamental Rights: A Commentary*, 2nd ed., Oxford Law Pro, 2024.
- Kelley, D.N., McCarthy, S.L., *The Report of the Crime Reporting Review Committee to Commissioner Raymond W. Kelly Concerning Compstat Auditing*, 2013.
- Kemelmacher-Shlizerman, I., Seitz, S.M., Miller, D., Brossard, E., *The Megaface Benchmark: 1 Million Faces for Recognition at Scale*, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, 4873-4882.
- Kemper, J., Kolkman, D., *Transparent to Whom? No Algorithmic Accountability Without a Critical Audience*, in *Information, Communication & Society*, 22(14), 2019, 2081-2096.
- Kent Police, *PredPol Operational Review – Initial Findings*, Restricted.
- Kerr, I., Earle, J., *Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy*, in *Stanford Law Review*, 66, 2013, 65-72.
- Kiela, D., Thrush, T., Ethayarajh, K., Singh, A., *Plotting Progress in AI*, Contextual AI Blog, 2023.
- Kindt, E.J., *Having Yes, Using No? About the New Legal Regime for Biometric Data*, in *Computer Law and Security Review*, 34(3), 2018, 523-538.
- Kitchin, R., *Big Data, New Epistemologies and Paradigm Shifts*, in *Big Data & Society*, 1(1), 2014, 1-12.
- Kleinberg, J. et al., *Discrimination in the Age of Algorithms*, NBER Working Paper Series No. 25548, 2019.

BIBLIOGRAPHY

- Kleinberg, J., Mullainathan, S., Raghavan, M., Inherent Trade-offs in the Fair Determination of Risk Scores, arXiv:1609.05807v2, 2016.
- Kluge Corrêa, N., Galvão, C., Santos, J.W., Del Pino, C., Pontes Pinto, E., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., De Oliveira, N., Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance, in *Patterns*, 4(10), 2023, Article 100857, 1-21.
- Koops, B.-J., The Concept of Function Creep, in *Law, Innovation and Technology*, 13(1), 2021, 29-56.
- Kosta, E., Boehm, F. (Eds.), *The EU Law Enforcement Directive (LED): A Commentary*, Oxford University Press, 2024.
- Kosta, E., Leenes, R., Kamara, I. (Eds.), *Research Handbook on EU Data Protection Law*, Edward Elgar Publishing, 2022.
- Kraemer, F., Van Overveld, K., Peterson, M., Is There an Ethics of Algorithms?, in *Ethics and Information Technology*, 13, 2011, 251-260.
- Krishnapriya, K.S., Vangara, K., King, M.C., Albiero, V., Bowyer, K., Characterizing the Variability in Face Recognition Accuracy Relative to Race, arXiv:1904.07325, 2019.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., ImageNet Classification with Deep Convolutional Neural Networks, in *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 2012, 1097-1105.
- Kuner, C., Cate, F., Lynskey, O., Millard, C., Ni Loideain, N., Svantesson, D., An Unstoppable Force and an Immoveable Object? EU Data Protection Law and National Security, *International Data Privacy Law*, 8(2), 2018, 151-167.
- Kurzweil, R., *The Age of Spiritual Machines*, Orion, 1999.
- Lagioia, F., Rovatti, R., Sartor, G., Algorithmic Fairness through Group Parities? The Case of COMPAS-SAPMOC, in *AI & Society*, 38, 2023, 459-478.
- Lagioia, F., Sartor, G., Il sistema COMPAS: algoritmi, previsioni, iniquità, in *XXVI Lezioni di Intelligenza Artificiale: Saggi a margine del ciclo seminariale "Intelligenza Artificiale e diritto"*, Giappichelli, 2020.

BIBLIOGRAPHY

- Lammerant, H., De Hert, P., Predictive Profiling and Its Legal Limits: Effectiveness Gone Forever, in B. Van der Sloot, D. Broeders, E. Schrijvers (Eds.), *Exploring the Boundaries of Big Data*, Amsterdam University Press, 2016, 145-173.
- Laney, D., 3D Data Management: Controlling Data Volume, Velocity and Variety, META Group Report, File 949, 2001.
- Laperruque, J., Facing the Future of Surveillance, The Constitution Project's Task Force on Facial Recognition Surveillance, Project on Government Oversight.
- Laux, J., Wachter, S., Mittelstadt, B., Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk, in *Regulation & Governance*, 18, 2024, 3-32.
- Lazar, S., Stone, J., On the Site of Predictive Justice, in *Noûs*, 58(3), 2024, 730-754.
- Lazcoz Moratinos, G., Human Oversight Or Monitoring In Article 14 Of The Artificial Intelligence Act: A Mere Mandatory Requirement For High-Risk Systems?, in L. Cotino Hueso, D.-U. Galetta (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, 2024, 717-737.
- Learned-Miller, E., Ordóñez, V., Morgenstern, J., Buolamwini, J., *Facial Recognition Technologies in the Wild: A Primer*, Algorithmic Justice League, 2020.
- Lechterman, T.M., The Concept of Accountability in AI Ethics and Governance, in J.B. Bullock et al. (Eds), *The Oxford Handbook of AI Governance*, Oxford University Press, 2022.
- Leibniz, G.W., *The New Method of Learning and Teaching Jurisprudence*, Lawbook Exchange Ltd, 2017.
- Leiser, M.R., Custers, B., The Law Enforcement Directive: Conceptual Challenges of EU Directive 2016/680, in *European Data Protection Law Review*, 5(3), 2019, 367-378.
- Lepri, B. et al., Fair, Transparent, and Accountable Algorithmic Decision-Making Processes, in *Philosophy & Technology*, 31(4), 2018, 611-627.
- Lequesne Roth, C. (Director), Caffin, J., Kimri, M., Kubiak, M., Lacour, C., Legros, P., Lemos, M., *La reconnaissance faciale dans l'espace – Une cartographie juridique européenne*, April 2020.
- Levitt, S.D., The Relationship between Crime Reporting and Police: Implications for the Use of Uniform Crime Reports, in *Journal of Quantitative Criminology*, 14(1), 1998, 61-81.

BIBLIOGRAPHY

- Li, J, Zhao, B, Zhang, H., Jiao, J., Face recognition system using SVM classifier and feature extraction by PCA and LDA combination, 2009 International Conference on Computational Intelligence and Software Engineering, Wuhan, China, 2009, 1-4.
- Li, L., Mu, X., Li, S., Peng, H., A Review of Face Recognition Technology, in IEEE Access, 8, 2020, 139110-139120.
- Li, X., Explainable AI (XAI) Empowering Criminal Psychological Profiling: In Pursuit of Transparency and Trust within the Black Box, in International Journal of Science, 12(7), 2025, 167-175.
- Liao, Q.V., Varshney, K.R., Human-Centered Explainable AI (XAI): From Algorithms to User Experiences, arXiv:2110.10790, 2021.
- Lindsay, R.K., Buchanan, B.G., Feigenbaum, E.A., Lederberg, J., DENDRAL: A Case Study of the First Expert System for Scientific Hypothesis Formation, in Artificial Intelligence, 61, 1993, 209-261.
- Lipton, Z.C., The Mythos of Model Interpretability, in Communications of the ACM, 61(10), 2018, 36-43.
- Liveley, G., Thomas, S., Homer's Intelligent Machines: AI in Antiquity, in S. Cave, K. Dihal, S. Dillon (Eds.), AI Narratives: A History of Imaginative Thinking about Intelligent Machines, Oxford University Press, 2020, 21-42.
- Lombardo, E., Sicurezza 4P: Lo studio alla base del software XLAW per prevedere e prevenire i crimini, Mazzanti Libri, 2019.
- Longo, E., Grounding Media Freedom in the EU: The Legal Basis of the EMFA, in Rivista italiana di informatica e diritto, 7(1), 2025, 111-124.
- Lum, K., Isaac, W., To Predict and Serve?, in Significance Magazine, 13(5), 2016, 14-19.
- Lum, K., Koper, C.S., Willis, J., Understanding the Limits of Technology's Impact on Police Effectiveness, in Police Quarterly, 20(2), 2017, 135–163.
- Lynskey, O., Criminal Justice Profiling and EU Data Protection Law: Precarious Protection from Predictive Policing, in International Journal of Law in Context, 15(2), 2019, 162-176.
- Lyon, D., La società sorvegliata: Tecnologie di controllo della vita quotidiana, Feltrinelli, 2003.

BIBLIOGRAPHY

- Madiega, T., Mildebrath, H., *Regulating Facial Recognition in the EU*, European Parliamentary Research Service, September 2021.
- Maguire, E., *Organizational Structure in American Police Agencies: Context, Complexity, and Control*, SUNY Press, 2003.
- Mahler, T., *Between Risk Management and Proportionality: The Risk-Based Approach in the EU's Artificial Intelligence Act Proposal*, in *Nordic Yearbook for ICT Law*, 2022, 227-254.
- Maira, G., *Intelligenza umana e intelligenza artificiale*, in *federalismi.it*, 7, 2021, V–XVIII.
- Maliszewska-Nienartowicz, J., *Direct and Indirect Discrimination in European Union Law – How to Draw a Dividing Line?*, in *International Journal of Social Sciences*, III(1), 2014, 41-55.
- Manning, P.K., *Information Technologies and the Police*, in *Modern Policing*, 15, 1992, 349-398.
- Manning, P.K., *Policing Contingencies*, The University of Chicago Press, 2003.
- Manning, P.K., *Technological Dramas and the Police: Statement and Counterstatement in Organizational Analysis*, in *Criminology*, 30(3), 1992, 327-346.
- Mansfield, T., *Facial Recognition Technology in Law Enforcement Equitability Study Test Strategy*, National Physical Laboratory, April 2022.
- Mantelero, A., *AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment*, in *Computer Law & Security Review*, 34(4), 2018, 754-772.
- Mantelero, A., *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI, IT&LAW* 36, Springer, 2022.
- Mantelero, A., *Comment to Articles 35 and 36*, in M. Cole, F. Boehm (Eds.), *GDPR Commentary*, Edward Elgar Publishing, 2019.
- Mantelero, A., Esposito, M.S., *An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems*, in *Computer Law & Security Review*, 41, 2021, Article 105561, 1-35.
- Mantelero, A., *The Artificial Intelligence Act: The European Legislator's Response To The Challenges Of Artificial Intelligence*, in L. Cotino Hueso, D.-U. Galetta (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, 2025, 45-51.

BIBLIOGRAPHY

- Mantelero, A., The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, Legal Obligations and Key Elements for a Model Template, in *Computer Law & Security Review*, 54, 2024, 1-18.
- Maple, J., Mitchell, C., *The Crime Fighter: How You Can Make Your Community Crime-Free*, Broadway Books, 2000.
- Marchetti, B., Amministrazione digitale, in *Enciclopedia del Diritto: funzioni amministrative*, Giuffrè, 2022, 75-109.
- Marchetti, B., Artificial Intelligence and Public Authorities: Does the European AI Act Protect Public Values?, in *European Review of Public Law*, 36(1), 2024, 67-90.
- Marchetti, B., Giustizia amministrativa e transizione digitale: Spunti per riflettere su un futuro non troppo lontano, in M. Ramajoli (Ed.), *Una giustizia amministrativa digitale?*, Il Mulino, 2023, 59-90.
- Marchetti, B., La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica, in *BioLaw Journal*, 2, 2021, 373-377.
- Marchetti, B., Torchia, L., AI and Public Administration, in F. Decarolis, B. Marchetti, L. Torchia (Eds.), *The EU Digital Regulation and its Impact on Member States*, Springer, 2025.
- Marquenie, T., The Police and Criminal Justice Authorities Directive: Data Protection Standards and Impact on the Legal Framework, in *Computer Law & Security Review*, 33(3), 2017, 324-340.
- Martín Delgado, I., Automación, inteligencia artificial y pública administración: vecchie categorie concettuali per nuovi problemi?, in *Le istituzioni del federalismo*, 3, 2019, 643-666.
- Martini, M., Regulating Artificial Intelligence - How to De-Mystify the Alchemy of Code?, in M. Ebers, S. Navas (Eds.), *Algorithms and Law*, Cambridge University Press, 2020, 100-135.
- Masso, A., Kasapoglu, T., Kaun, A., Galis, V., Citizens' Perspectives on Platformisation of Police Work: A Scenario and Story-Based Exploration in Estonia and Sweden, in *Information, Communication & Society*, 27(13), 2024, 2400-2418.
- Mastrobuoni, G., Crime Is Terribly Revealing: Information Technology and Police Productivity, in *The Review of Economic Studies*, 87(6), 2020, 2727-2753.

BIBLIOGRAPHY

- Mastrofski, S.D., Willis, J.T., Police Organization Continuity and Change: Into the Twenty-First Century, in *Crime and Justice*, 39(1), 2010, 55-144.
- Mattarella, B.G., Discrezionalità amministrativa, in S. Cassese (Ed.), *Dizionario di diritto pubblico*, III, Giuffrè, Milano, 2006.
- Mayer-Schönberger, V., Cukier, K., *Big Data: A Revolution That Will Transform How We Live, Work and Think*, Houghton Mifflin Harcourt, 2013.
- Mayson, S., Bias In, Bias Out, University of Georgia School of Law Legal Studies Research Paper, No. 2018-35, 2018.
- Mayson, S.G., Bias In, Bias Out, *The Yale Law Journal*, 128(8), 2019, 2218-2298.
- McCarthy, J., Feigenbaum, E., In Memoriam, *AI Magazine*, 11(3), 1990, 10-11.
- McCarthy, J., Minsky, M., Shannon, C., Rochester, N., A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, 1955.
- McCarthy, J., Programs with Common Sense, Computer Science Department, Stanford University, 1959.
- McCarthy, J., What Is Artificial Intelligence?, Stanford University, 2007.
- McCorduck, P., *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*, Routledge, 2004.
- McCulloch, J., Wilson, D., *Pre-crime Pre-emption, Precaution and the Future*, Routledge, 2015.
- McCulloch, W.S., Pitts, W., A Logical Calculus of Ideas Immanent in Nervous Activity, in *Bulletin of Mathematical Biophysics*, 5, 1943, 115-133.
- McDermott, J., R1: A Rule-Based Configurer of Computer Systems, in *Artificial Intelligence*, 19(1), 1982, 39-88.
- McFadden, M. et al. (Oxford Commission on AI & Good Governance), *Harmonising Artificial Intelligence: The Role of Standards in the EU AI Regulation*, December 2021.
- Merler, M., Il ruolo della Commissione Europea nella realizzazione dello spazio digitale europeo: tra costituzionalismo digitale e sovranità digitale, in *Rivista trimestrale di diritto pubblico*, 4, 2024, 1121-1137.

BIBLIOGRAPHY

- Merler, M., La Corte costituzionale tedesca si pronuncia sul trattamento automatizzato dei dati in contesti di polizia, in *Giornale di diritto amministrativo*, 5, 2023, 685-694.
- Merler, M., Ramotti, C., Italian Public Administration in the Face of Digital Transformation, in F. Decarolis, B. Marchetti, L. Torchia (Eds.), *The EU Digital Regulation and its Impact on Member States*, Springer, 2025, 35-76.
- Merler, M., Ratha, N., Feris, R.S., Smith, J.R., Diversity in Faces, arXiv:1901.10436v6, 2019.
- Metikoš, L., Ausloos, J., The Right to an Explanation in Practice: Insights from Case Law for the GDPR and the AI Act, in *Law, Innovation and Technology*, 2025, 1-36.
- Migliorini, S., Il ravvicinamento delle legislazioni nella prospettiva del diritto dell'Unione europea e dei suoi effetti sul diritto interno. Le tecniche del ravvicinamento, in *Osservatorio sulle fonti*, 3, 2012, 1-28.
- Min Chen, Shiwen Mao, Yunhao Liu, Big Data: A Survey, in *Mobile Networks and Applications*, 19, 2014, 171-209.
- Ming-Hsuan, Y., Kriegman, D.J., Ahuja, N., Detecting Faces in Images: A Survey, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1), 2002, 34-58.
- Minsky, M., Papert, S., *Perceptrons: An Introduction to Computational Geometry*, MIT Press, 1969.
- Minsky, M., Steps toward Artificial Intelligence, in *Proceedings of the IRE*, 49(1), 1961, 8-30.
- Miró Llinares, F., Santisteban Galarza, M., The Regulation of Predictive Policing Systems in the Artificial Intelligence Act, in L. Cotino Hueso, D.-U. Galetta (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, Napoli, 2024, 340-349.
- Mitchell, M., *L'intelligenza artificiale: Una guida per esseri umani pensanti*, Einaudi, Torino, 2022.
- Mitchell, S., Potash, E., Barocas, S., D'Amour, A., Lum, K., Algorithmic Fairness: Choices, Assumptions, and Definitions, in *Annual Review of Statistics and Its Application*, 8, 2021, 141-163.
- Mitchell, T.M., *Machine Learning*, McGraw-Hill, Boston, 1997.
- Mitsilegas, V., Autonomous Concepts, Diversity Management and Mutual Trust in Europe's Area of Criminal Justice, in *Common Market Law Review*, 57(1), 2020, 45-78.

BIBLIOGRAPHY

- Mitsilegas, V., The Transformation of Privacy in an Era of Pre-emptive Surveillance, in *Tilburg Law Review*, 20(1), 2015, 35-57.
- Mittelstadt, B., Russell, C., Wachter, S., Explaining Explanations in AI, in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, 2019, 279-288.
- Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L., The Ethics of Algorithms: Mapping the Debate, in *Big Data & Society*, 3(2), 2016, 1-21.
- Mobilio, G., La Corte EDU condanna il ricorso alle tecnologie di riconoscimento facciale per reprimere il dissenso politico: osservazioni a partire dal caso Glukhin c. Russia, in *DPCE Online*, 62(1), 2024, 695-705.
- Mobilio, G., L'intelligenza artificiale e le regole giuridiche alla prova: il caso paradigmatico del GDPR, *Federalismi.it*, 16, 2020, 248-290.
- Mobilio, G., Tecnologie di riconoscimento facciale: Rischi per i diritti fondamentali e sfide regolatorie, *Editoriale Scientifica*, 2021.
- Mobilio, G., Your Face Is Not New to Me – Regulating the Surveillance Power of Facial Recognition Technologies, in *Internet Policy Review*, 12(1), 2023, 1–31.
- Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P., Tita, G.E., Self-Exciting Point Process Modeling of Crime, in *Journal of the American Statistical Association*, 106(493), 2011, 100-108.
- Molesworth, W., *The Collected Works of Thomas Hobbes*, Routledge, 1997.
- Molnar, C., *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed., Independently Published, 2020.
- Molnár, T., Algorithmic Decision-Making in Migration and Security in the EU: Challenges in Ensuring Effective Legal Remedies, in M. Varju, K. Mezei (Eds.), *The Challenges of Artificial Intelligence for Law in Europe. CAILEIC 2022. Data Science, Machine Intelligence, and Law*, Vol. 6, Springer, Cham, 2023, 147-172.
- Monika, M., Rajender, U., Tamizhselvi, A., Rumale, A.S., Real-Time Object Detection in Videos Using Deep Learning Models, in *ICTACT Journal on Image and Video Processing*, 14(2), 2023, 3103-3109.

BIBLIOGRAPHY

- Montag, L. et al., The Rise and Rise of Biometrics Mass Surveillance in the EU: A Legal Analysis of Biometrics Mass Surveillance Practices in Germany, the Netherlands, and Poland, EDRi – European Digital Rights, 2021.
- Moor, J., The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years, in *AI Magazine*, 27(4), 2006, 87-91.
- Mordini, E., Tzovaras, D. (Eds.), *Second Generation Biometrics: The Ethical, Legal and Social Context*, Springer, 2012.
- Morley, J., Floridi, L., Kinsey, L., Elhalal, A., From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices, in *Science and Engineering Ethics*, 26, 2020, 2141-2168.
- Moses, L.B., Chan, J., Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability, in *Policing and Society*, 28(7), 2018, 806-822.
- Moses, Y.A., Ullman, S., Face Recognition: The Problem of Compensating for Changes in Illumination Direction, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 1997, 721-732.
- Mosher, C.J., Miethe, T.D., Hart, T.C., *The Mismeasure of Crime*, Sage Publications, Thousand Oaks, 2010.
- Moussawi, S., Deng, X., Joshi, K.D., AI and Discrimination: Sources of Algorithmic Biases, in *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, 55(4), 2024, 6-11.
- Murdick, D., Dunham, J., Melot, J., AI Definitions Affect Policymaking, Center for Security and Emerging Technology, June 2020.
- Nakar, S., Greenbaum, D., Now You See Me. Now You Still Do: Facial Recognition Technology and the Growing Lack of Privacy, in *Journal of Science & Technology Law*, 23(1), 2017, 88-122.
- Nannini, L., Habemus a Right to an Explanation: So What? A Framework on Transparency-Explainability Functionality and Tensions in the EU AI Act, in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 7(1), 2024, 1023-1035.
- Nardocci, C., Artificial Intelligence-based Discrimination: Theoretical and Normative Responses. Perspectives from Europe, *DPCE Online*, 60(3), 2023, 2367–2394.

BIBLIOGRAPHY

- National Academies of Science, Engineering and Medicine, Facial Recognition Technology: Current Capabilities, Future Prospects, and Governance, The National Academies Press, Washington, DC, 2024.
- National Institute of Justice, Predictive Policing Symposiums.
- Neerukonda, M., Chaudhuri, B., Are Technologies (Gender)-Neutral?: Politics and Policies of Digital Technologies, in *ASCI Journal of Management*, 47, 2018, 32-44.
- Neroni Rezende, I., Facial Recognition in Police Hands: Assessing the 'Clearview Case' from a European Perspective, in *New Journal of European Criminal Law*, 11(3), 2020, 375-389.
- Newburn, T., *Handbook of Policing*, Willan Publishing, 2003.
- Newell, A., Intellectual Issues in the History of Artificial Intelligence, in F. Machlup, U. Mansfield (Eds.), *The Study of Information: Interdisciplinary Messages*, John Wiley & Sons, 1983, 187-294.
- Newell, A., Simon, H.A., Computer Science as Empirical Inquiry: Symbols and Search, in *Communications of the ACM*, 19(3), 1976, 113-126.
- Ngan, M., Grother, P., Face Recognition Vendor Test (FRVT): Performance of Automated Gender Classification Algorithms, NIST, April 2015.
- Nilsson, N.J., *Artificial Intelligence: A New Synthesis*, Morgan Kaufmann Publishers, 1998.
- NIST, Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information, NIST Special Publication 500-290, January 2011.
- NIST, Four Principles of Explainable Artificial Intelligence, NISTIR 8312, September 2021.
- Northrop, A., Kraemer, K.L., King, J.L., Police Use of Computers, in *Journal of Criminal Justice*, 23(3), 1995, 259-275.
- Norval, C., Cornelius, K., Cobbe, J., Singh, J., Disclosure by Design: Designing Information Disclosures to Support Meaningful Transparency and Accountability, in 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21-24, 2022, Seoul, Republic of Korea, ACM, New York, 2022, 679-690.
- Novelli, C., Casolari, F., Rotolo, A., Taddeo, M., Floridi, L., Taking AI Risks Seriously: A New Assessment Model for the AI Act, in *AI & SOCIETY*, 39, 2023, 1641-1660.

BIBLIOGRAPHY

- Nowotny, H., In *AI We Trust: Power, Illusion and Control of Predictive Algorithms*, Polity Press, 2021.
- Nunn, S., *Police Information Technology: Assessing the Effects of Computerization on Urban Police Functions*, in *Public Administration Review*, 61(2), 2001, 221-234.
- Nyman Metcalf, K., Papageorgiou, I., *The European Union Digital Single Market—Challenges and Impact for the EU Neighbourhood States*, in *Baltic Journal of European Studies*, 8(2), 2018, 7-23.
- Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S., *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, in *Science*, 366, 2019, 447-453.
- OECD, *Big Data: Bringing Competition Policy to the Digital Era*, 2016.
- OECD, *Explanatory Memorandum on the Updated OECD Definition of an AI System*, OECD Artificial Intelligence Papers, No. 8, OECD Publishing, Paris, 2024.
- OECD, *Framework for the Classification of AI Systems*, OECD Digital Economy Papers, No. 323, OECD Publishing, Paris, 2022.
- OECD, *Glossary of Statistical Terms*, 2008.
- OECD, *Measuring the Environmental Impacts of Artificial Intelligence Compute and Applications*, OECD Digital Economy Papers, No. 341, OECD Publishing, 2022.
- OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, May 2019 (revised May 2024).
- OECD, *Scoping the OECD AI Principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)*, OECD Digital Economy Papers, No. 291, OECD Publishing, 2019.
- OECD, *Updates to the OECD's Definition of an AI System Explained*, November 2023.
- O'Flaherty, M., *Opinions, Facial Recognition Technology and Fundamental Rights*, in *European Data Protection Law Review*, 6(2), 2020, 170-173.
- Ogg, J., *Preventive Justice and the Power of Policy Transfer*, Springer, 2015.
- Olivato, G., *The AI Act: Regulatory and Technical Challenges*, Doctoral Thesis, University of Trento, 20 June 2025.

BIBLIOGRAPHY

- Onitui, D., The Limits of Explainability & Human Oversight in the EU Commission's Proposal for the Regulation on AI - A Critical Approach Focusing on Medical Diagnostic Systems, in *Information & Communications Technology Law*, 32(2), 2023, 174-194.
- Oosterloo, S., Van Schie, G., The Politics and Biases of the “Crime Anticipation System” of the Dutch Police, *Proceedings of the International Workshop on Bias in Information, Algorithms, and Systems*, 2018, 2103, 30-41.
- Orlandi, R., Uso poliziesco dell'intelligenza artificiale. L'insegnamento del Bundersverfassungsgericht, in *Cassazione Penale*, 7/8, 2023, 2167-2180.
- Orofino, A.G., La patologia dell'atto amministrativo elettronico: sindacato giurisdizionale e strumenti di tutela, in *Foro Amministrativo-C.d.S.*, 9, 2022, 2256-2281.
- Orsoni, G., D'Orlando, E., Nuove prospettive dell'amministrazione digitale: Open Data e algoritmi, in *Le istituzioni del federalismo*, 3, 2019, 593-617.
- Ortega Giménez, A., The Territorial Scope Of Application Of The Artificial Intelligence Act, in L. Cotino Hueso, D.-U. Galetta (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, 2025, 115-129.
- Ortigosa, A.P., Conformity Assessment In The Design And Production Of Artificial Intelligence-Based Systems In The Context Of The "New Legislative Framework", in L. Cotino Hueso, D.-U. Galetta (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, 2025, 449-470.
- Ortigosa, A.P., General Regime of Obligations for Providers and Deployers in the Artificial Intelligence Act, in L. Cotino Hueso, D.-U. Galetta (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, 2025, 471-522.
- OSCE, *Guidebook Intelligence-Led Policing*, 2017.
- Oswald, M., Babuta, A., *Data Analytics and Algorithmic Bias in Policing*, Royal United Services Institute for Defence and Security Studies, 2019.
- Oswald, M., Grace, J., Intelligence, policing and the use of algorithmic analysis: a freedom of information-based study, in *Journal of Information Rights, Policy & Practice*, 1(1), 2016, 1-14.
- Oswald, M., Grace, J., Urwin, S., Barnes, G.C., Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and 'Experimental' Proportionality, in *Information & Communications Technology Law*, 27(2), 2018, 223-250.

BIBLIOGRAPHY

- Palazzani, L., *Information Technologies and Artificial Intelligence: Sfide etiche al diritto*, Edizioni Studium, 2020.
- Pangelinan, G., Krishnapriya, K.S., Albiero, V., Bezold, G., Zhang, K., Vangara, K., King, M.C., Bowyer, K.W., *Exploring Causes of Demographic Variations in Face Recognition Accuracy*, arXiv:2304.07175v1, 2023.
- Papakonstantinou, V., De Hert, P., *The Regulation of Digital Technologies in the EU: The Law-Making Phenomena of 'Act-ification', 'GDPR Mimesis' and 'EU Law Brutality'*, in *Technology and Regulation Journal*, 2022, 50-62.
- Parkhi, O.M., Vedaldi, A., Zisserman, A. et al., *Deep Face Recognition*, *Proceedings of the British Machine Vision Conference 2015*, 1–12.
- Parona, L., *Prospettive europee e internazionali di regolazione dell'intelligenza artificiale tra principi etici, soft law e self-regulation*, in *Rivista della Regolazione dei mercati*, 1, 2020, 70-93.
- Pascal, B., *Pascal's Pensées*, E.P. Dutton & Co., Inc., New York, 1958.
- Pasquale, F., *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, 2015.
- Pasquino, T., *Identità digitale della persona, diritto all'immagine e reputazione*, in E. Tosi (Ed.), *Privacy digitale: Riservatezza e protezione dei dati personali tra GDPR e nuovo Codice Privacy*, Giuffrè, Milano, 2019, 93-114.
- Patroni Griffi, F. (Ed.), *La Trasparenza della pubblica amministrazione tra accessibilità totale e riservatezza*, in *federalismi.it*, 8, 2013.
- Patterson, D., Gonzalez, J., Holzle, U., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D.R., Texier, M., Dean, J., *The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink*, in *Computer*, 55(7), 2022, 18-28.
- Paul, R., *European Artificial Intelligence "Trusted Throughout the World": Risk-Based Regulation and the Fashioning of a Competitive Common AI Market*, in *Regulation and Governance*, 18(4), 2024, 1067-1085.
- Pearl, J., *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann Publishers, 1988.
- Pearsall, B., *Predictive Policing: The Future of Law Enforcement?*, in *National Institute of Justice Journal*, 266, 2010, 16-19.

BIBLIOGRAPHY

- Penev, P.S., Atick, J.J., Local Feature Analysis: A General Statistical Theory for Object Representation, in *Network: Computation in Neural Systems*, 7(3), 1996, 477-500.
- Perry, W.L., McInnis, B., Price, C.C., Smith, S.C., Hollywood, J.S., Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations, RAND Corporation, 2013.
- Peters, T.M., Visser, R.W., The Importance of Distrust in AI, in *Explainable Artificial Intelligence First World Conference, xAI 2023*, Lisbon, Portugal, July 26-28, 2023, *Proceedings, Part III* (CCIS, volume 1903), arXiv:2307.13601, 2023.
- Phillips, P., Moon, H., Rizvi, S., Rauss, P., The FERET Evaluation Methodology for Face-Recognition Algorithms, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 2000, 1090-1104.
- Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W., Overview of the Face Recognition Grand Challenge, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 1, 2005, 947-954.
- Pietrocarlo, E., Predictive Policing: criticità e prospettive dei sistemi di identificazione dei potenziali criminali, in *Diritto Penale Contemporaneo*, 2, 2023, 114-158.
- Piras, A., Discrezionalità amministrativa, in *Enciclopedia del Diritto*, Volume XIII, Giuffrè, 1964.
- Pisati, M., Indagini preliminari e intelligenza artificiale: efficienza e rischi per i diritti fondamentali, in *Processo penale e giustizia*, 4, 2020, 957-971.
- Pitch, T., *La società della prevenzione*, Carocci, 2007.
- Pizzetti, F. (Ed.), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Giappichelli, 2018.
- Pizzetti, F., La protezione dei dati personali e la sfida dell'Intelligenza Artificiale, in F. Pizzetti (Ed.), *Intelligenza Artificiale, protezione dei dati personali e regolazione*, Giappichelli, 2018, 61-120.
- Police Executive Research Forum, *Excellence in Problem-Oriented Policing: The 2001 Herman Goldstein Award Winners*, 2011.
- Police Leaders at PERF/BJA Meeting Discuss CompStat: Best Practices and Future Outlook, in *Newsletter of the Police Executive Research Forum*, 25(2), March/April 2011.
- Postman, N., *Technopoly: The Surrender of Culture to Technology*, Vintage Books, 1992.

BIBLIOGRAPHY

- Pouget, H., Zuhdi, R., AI and Product Safety Standards Under the EU AI Act, Carnegie Endowment, 5 March 2024.
- President's Commission on Law Enforcement and Administration of Justice, The Challenge of Crime in a Free Society, Report, 1967.
- Presno Linera, M.Á., The Prohibition Of Artificial Intelligence Systems That Evaluate And Classify People Based On Data That Are Unrelated To The Context In Which They Were Generated And That Lead To Discrimination, in L. Cotino Hueso, D.-U. Galetta (Eds.), The European Union Artificial Intelligence Act. A Systematic Commentary, CERIDAP, Editoriale Scientifica, 2025, 241-255.
- Pressacco, L., Intelligenza artificiale e ragionamento probatorio nel processo penale, in BioLaw Journal, 4, 2022, 503-530.
- Prifti, K., Morley, J., Novelli, C., Floridi, L., Regulation by Design: Features, Practices, Limitations, and Governance Implications, in Minds and Machines, 34, 2024, Article 13, 1-23.
- Prifti, K., Regulating Artificial Intelligence by Design: Exploring the Practice, Developing the Theory, and Devising Governance Solutions, Doctoral Thesis, Erasmus University Rotterdam, 2024.
- Qinjun, L., Tianwei, C., Yan, Z., Yuying, W., Facial Recognition Technology: A Comprehensive Overview, in Academic Journal of Computing & Information Science, 6(7), 2023, 15-26.
- Quattrocolo, S., Artificial Intelligence, Computational Modelling and Criminal Proceedings: A Framework for A European Legal Discussion, Springer, Legal Studies in International, European and Comparative Criminal Law, Cham, 2020.
- Quezada Tavárez, K., Impact of the Right of Access on the Balance between Security and Fundamental Rights: Informational Power as a Tool to Watch the Watchers, in European Data Protection Law Review, 7(1), 2021, 59-73.
- Quintel, T., Data Protection Rules Applicable to Financial Intelligence Units: Still No Clarity in Sight, in ERA Forum, 23, 2022, 53-74.
- Quintel, T., Removing Obstacles for Cooperation between EU Financial Intelligence Units: The Data Protection Directive for Police and Criminal Justice Authorities as Fitting Data Protection Instrument, May 2019.

BIBLIOGRAPHY

- Rachovitsa, A., Johann, N., The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case, in *Human Rights Law Review*, 22(2), 2022, 1-15.
- Radiya-Dixit, E., A Sociotechnical Audit: Assessing Police Use of Facial Recognition, Minderoo Centre for Technology and Democracy, University of Cambridge, October 2022.
- Raji, I.D., Fried, G., About Face: A Survey of Facial Recognition Evaluation, arXiv:2102.00813, 2021.
- Ramón Fernández, F., Quality Management Systems, Technical Documentation And Documentation Keeping In The Regulation, in L. Cotino Hueso, D.-U. Galetta (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, 2024, 629-647.
- Rangone, N., Intelligenza artificiale, tutela dell'ambiente e regolazione europea. *Biolaw Journal*, *Rivista di BioDiritto*, 1, 2025, 529-548.
- Rangone, N., Megale, L., Risks Without Rights? The EU AI Act's Approach to AI in Law and Rule-making . *European Journal of Risk Regulation*, 16(3), 2025, 1082-1097.
- Rashid, B., Kausik, A.K., AI Revolutionizing Industries Worldwide: A Comprehensive Overview of Its Diverse Applications, in *Hybrid Advances*, 7, 2024, Article 100277.
- Ratcliffe, J., Advocate: Predictive Policing, in D. Weisburd, A.A. Braga (Eds.), *Police Innovation: Contrasting Perspectives*, Cambridge University Press, 2019, 347-365.
- Ratcliffe, J.H., *Intelligence-Led Policing*, Routledge, 2016.
- Ratledge, E.C., Jacoby, J.E., *Handbook on Artificial Intelligence and Expert Systems in Law Enforcement*, Greenwood, 1989.
- Reidenberg, J.R., Privacy in Public, in *University of Miami Law Review*, 69(1), 2014, 141-159.
- Renan, D., The Fourth Amendment as Administrative Governance, in *Stanford Law Review*, 68, 2016, 1039-1130.
- Renda, A., Study to Support an Impact Assessment of Regulatory Requirements for Artificial Intelligence in Europe, Final Report (D5), April 2021.
- Renda, A., Engler, A., What's in a Name? Getting the Definition of Artificial Intelligence Right in the EU's AI Act, CEPS Explainers, 2023.

BIBLIOGRAPHY

- Resta, G., Identità personale e identità digitale, in *Il Diritto dell'informazione e dell'informatica*, 2007, 511-531.
- Reviglio, U., The Untamed and Discreet Role of Data Brokers in Surveillance Capitalism: A Transnational and Interdisciplinary Overview, in *Internet Policy Review*, 11(3), 2022, 1-27.
- Richardson, R., Defining and Demystifying Automated Decision Systems, in *Maryland Law Review*, 81(3), 2022, 785-840.
- Richardson, R., Schultz, J.M., Crawford, K., Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice, in *New York University Law Review Online*, 94, 2019, 15-55.
- Rijpma, J., Et al. (Eds), *The New EU Data Protection Regime: Setting Global Standards for the Right to Personal Data Protection*, the XXIX FIDE Congress in The Hague 2020 Congress Publications, vol. 2, Eleven International Publishing, the Hague, 2020.
- Ripellino, A.M., *Magic Prague*, University of California Press, 1993.
- Rivling, G., *AI Valley: Microsoft, Google, and the Trillion-Dollar Race to Cash In on Artificial Intelligence*, Harper Business, 2025.
- Roig, A., Safeguards for the Right Not to Be Subject to a Decision Based Solely on Automated Processing (Article 22 GDPR), in *European Journal of Law and Technology*, 8(3), 2017, 1-17.
- Rose-Ackerman, S., Billa, B., Treaties and National Security, in *International Law and Politics*, 40, 2008, 437-496.
- Rosenblatt, F., The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, in *Psychological Review*, 65(6), 1958, 386-408.
- Rosenblueth, A., Wiener, N., Bigelow, J., Behavior, Purpose and Teleology, in *Philosophy of Science*, 10(1), 1943, 18-24.
- Rowe, E.A., Regulating Facial Recognition Technology in the Private Sector, in *Stanford Technology Law Review*, 24(1), 2021, 1-54.
- Roy, K.E., Defrosting the Chill: How Facial Recognition Technology Threatens Free Speech, in *Roger Williams University Law Review*, 27(1), 2022, 185-210.
- RTI International Police Executive Research Forum, *Research on the Impact of Technology on Policing Strategy in the 21st Century*, May 2016.

BIBLIOGRAPHY

- Rubinstein, I.S., Big Data: The End of Privacy or a New Beginning?, in *International Data Privacy Law*, 3(2), 2013, 74-87.
- Rudin, C., Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead, in *Nature Machine Intelligence*, 1, 2019, 206-215.
- Rudin, C., Wang, C., Coker, B., The Age of Secrecy and Unfairness in Recidivism Prediction, in *Harvard Data Science Review*, 2(1), 2020, 1-53.
- Rudschies, C., Schneider, I., The Long and Winding Road to Bans for Artificial Intelligence: From Public Pressure and Regulatory Initiatives to the EU AI Act, in *Digital Society*, 4, 2025, Article 57.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., Learning Representations by Back-Propagating Errors, in *Nature*, 323(6088), 1986, 533-536.
- Rumelhart, D.E., McClelland, J.L. (Eds.), *Parallel Distributed Processing*, MIT Press, Cambridge (US), 1986.
- Russell, S., Norvig, P., *Artificial Intelligence: A Modern Approach*, Pearson Education Inc., 2010.
- Russell, S., Rise of AI Debaters Highlights Need for Transparency, in *Nature*, 592, 2021, 166.
- Sachoulidou, A., Harnessing AI for Law Enforcement: Solutions and Boundaries from the Forthcoming AI Act, in *New Journal of European Criminal Law*, 15(2), 2024, 122-145.
- Sajfert, J., Quintel, T., *Data Protection Directive (EU) 2016/680 for Police and Criminal Justice Authorities*, December 2017.
- Salazar, I., Liébanas, M.Á., Post-Market Monitoring On High-Risk AI Systems In The Artificial Intelligence Act. Description, Measures And Use Cases, in L. Cotino Hueso, D.-U. Galetta (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, Napoli, 2024, 785-797.
- Salvador Serna, M., Inteligencia artificial y gobernanza de datos en la Administración Pública: sentando las bases para su integración a nivel corporativo, in *Repensando la administración pública: administración digital e innovación pública*, INAP, 2021, 126-148.
- Samoili, S., López Cobo, M., Delipetrev, B., Martínez-Plumed, F., Gómez, E., De Prato, G., *AI Watch: Defining Artificial Intelligence 2.0 – Towards an Operational Definition and Taxonomy*

BIBLIOGRAPHY

for the AI Landscape, EUR 30873 EN, Publications Office of the European Union, Luxembourg, 2021.

Samoili, S., Lopez Cobo, M., Gomez Gutierrez, E., De Prato, G., Martinez-Plumed, F., Delipetrev, B., AI Watch: Defining Artificial Intelligence, EUR 30117 EN, Publications Office of the European Union, Luxembourg, 2020.

Samuel, A.L., Some Studies in Machine Learning Using the Game of Checkers, in IBM Journal of Research and Development, 3(3), 1959, 210-229.

Santosuosso, A., Forum AI and Law, in BioLaw Journal, 1, 2020, 469-475.

Santosuosso, A., Sartor, G., Decidere con l'IA. Intelligenze artificiali e naturali nel diritto, Il Mulino, Bologna, 2024.

Sarker, I.H., Machine Learning: Algorithms, Real-World Applications and Research Directions, in SN Computer Science, 2, 2021, Article 160, 1-21.

Sartor, G., Lagioia, F., The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence, Study for the European Parliament, 2020.

Sartor, G., L'informatica giuridica e le tecnologie dell'informazione: Corso d'informatica giuridica, Giappichelli, Torino, 2022.

Sartor, G., Loreggia, A., The Impact of Algorithms for Online Content Filtering or Moderation. Upload Filters, Study commissioned by the European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs, 2020.

Scantamburlo, T., Charlesworth, A., Cristianini, N., Machine Decisions and Human Consequences, in K. Yeung, M. Lodge (Eds.), Algorithmic Regulation, Oxford University Press, 2019, 49-81.

Schapel, H., The New Approach to the New Approach: The Juridification of Harmonized Standards in EU Law, in Maastricht Journal of European and Comparative Law, 20(4), 2013, 521-533.

Scherer, M.U., Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies, in Harvard Journal of Law & Technology, 29(2), 2016, 354-400.

Schmidt, R.M., Recurrent Neural Networks (RNNs): A Gentle Introduction and Overview, arXiv:1912.05911.

Schneeberger, D. et al., The Tower of Babel in Explainable Artificial Intelligence (XAI), in A. Holzinger, P. Kieseberg, F. Cabitza, A. Campagner, A.M. Tjoa, E. Weippl (Eds.), Machine

BIBLIOGRAPHY

- Learning and Knowledge Extraction, CD-MAKE 2023, Lecture Notes in Computer Science, Volume 14065, Springer, Cham, 2023, 73-81.
- Schroff, S., The Purpose of Copyright – Moving Beyond the Theory, in *Journal of Intellectual Property Law & Practice*, 16(11), 2021, 1262-1272.
- Schuett, J., Defining the Scope of AI Regulations, in *Law, Innovation and Technology*, 15(1), 2023, 60-82.
- Schuett, J., Risk Management in the Artificial Intelligence Act, in *European Journal of Risk Regulation*, 15, 2024, 367-385.
- Schuilenburg, M., Soudijn, M., Big Data Policing: The Use of Big Data and Algorithms by the Netherlands Police, in *Policing: A Journal of Policy and Practice*, 17, 2023, Article paad061.
- Scientific Foresight Unit (STOA), Artificial Intelligence: From Ethics to Policy, European Parliament, 2020.
- Scoca, F.G., Attività amministrativa, in *Enc. Dir., Agg., VI*, Giuffrè, Milano, 2002.
- Scoca, F.G., La discrezionalità nel pensiero di Giannini e nella dottrina successiva, in *Rivista Trimestrale di Diritto Pubblico*, 2000, 1045-1078.
- Searle, J., Minds, Brains, and Programs, in *Behavioral and Brain Sciences*, 3(3), 1980, 417-424.
- Selinger, E., Hartzog, W., The Inconsentability of Facial Surveillance, in *Loyola Law Review*, 66, 2019, 101-122.
- Selmini, R., Sicurezza urbana e prevenzione della criminalità in Europa: alcune riflessioni comparate, in *Polis, Ricerche e studi su società e politica*, 1, 1999, 69-76.
- Senden, L.A.J., Soft Law and Its Implications for Institutional Balance in the EC, in *Utrecht Law Review*, IGITUR, 1(2), 2005, 79-99.
- Senden, L.A.J., Soft Law, Self Regulation and Co-regulation in European Law: Where Do They Meet, in *Electronic Journal of Comparative Law*, 9(1), 2005, 1-27.
- Sgaier, S.K., Huang, V., Charles, G., The Case for Causal AI, in *Stanford Social Innovation Review*, 18(3), 2020, 50-55.

BIBLIOGRAPHY

- Shahbazi, N., Lin, Y., Asudeh, A., Jagadish, H.V., Representation Bias in Data: A Survey on Identification and Resolution Techniques, in *ACM Computing Surveys*, 55(13), 2023, Article No. 293, 1-39.
- Sharif, M., Naz, F., Yasmin, M., Shahid, M.A., Rehman, A., Face Recognition: A Survey, in *Journal of Engineering Science & Technology Review*, 10(2), 2017, 166-177.
- Shearer, C., The CRISP-DM Model: The New Blueprint for Data Mining, in *Journal of Data Warehousing*, 5(4), 2000, 13-22.
- Sheikh, H., Prins, C., Schrijvers, E., *Mission AI: The New System Technology*, Springer, 2023.
- Shelley, M.W., *Frankenstein or, The Modern Prometheus*, Modern Library, 1993.
- Sherman, L., Gartin, P., Buerger, M., Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place, in *Criminology*, 27(1), 1989, 27-56.
- Sherman, L.W., Evidence-Based Policing, in *Ideas in American Policing*, 1998.
- Sherman, L.W., Neyroud, P.W., *Offender-Desistance Policing and the Sword of Damocles*, Civitas: Institute for the Study of Civil Society, London, 2012.
- Sherman, L.W., Offender-Desistance Policing (ODP): Less Prison and More Evidence in Rehabilitating Offenders, in T. Bliesener, A. Beelmann, M. Stemmler (Eds.), *Antisocial Behavior and Crime: Contributions of Developmental and Evaluation Research to Prevention and Intervention*, Hogrefe Publishing, 2012, 199-218.
- Shortliffe, E.H., Buchanan, B.G., A Model of Inexact Reasoning in Medicine, in *Mathematical Biosciences*, 23, 1975, 351-379.
- Simoncini, A., Amministrazione digitale algoritmica. Il Quadro costituzionale, in R. Cavallo Perin, D.-U. Galetta (Eds.), *Il Diritto Dell'amministrazione Pubblica Digitale*, Giappichelli, 2020.
- Simoncini, A., L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà, in *BioLaw Journal*, 1, 2019, 63-89.
- Simoncini, A., Profili costituzionali dell'amministrazione algoritmica, in *Rivista Trimestrale di Diritto Pubblico*, 4, 2019, 1149-1188.
- Simoncini, A., Verso la regolamentazione della Intelligenza Artificiale. Dimensioni e governo, in *BioLaw Journal*, 2, 2021, 414-417.

BIBLIOGRAPHY

- Skitka, L.J., Mosier, K.L., Burdick, M., Does Automation Bias Decision Making?, in *International Journal of Human-Computer Studies*, 51(5), 1999, 991-1006.
- Slobogin, C., Administrative Regulation of Programmatic Policing: Why "Leaders of a Beautiful Struggle" is Both Right and Wrong, in *Administrative Law Review*, 75(3), 2023, 565-586.
- Slobogin, C., Government Dragnets, in *Law and Contemporary Problems*, 73(3), 2010, 101-132.
- Slobogin, C., Panvasive Surveillance, Political Process Theory and the Nondelegation Doctrine, in *Georgetown Law Journal*, 102, 2014, 1721–1776.
- Smuha, N.A., Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea, in *Philosophy & Technology*, 34, 2021, 91-104.
- Smuha, N.A., Beyond the individual: Governing AI's Societal Harm, in *Internet Policy Review*, 10(3), 2021, 1-32.
- Smuha, N.A., Yeung, K., The European Union's AI Act: Beyond Motherhood and Apple Pie?, in N.A. Smuha (Ed.), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*, Cambridge Law Handbooks, Cambridge University Press, 2025, 228-258.
- Soler Garrido, J., De Nigris, S., Bassani, E., Sánchez, I., Evas, T., André, A.-A., Boulangé, T., Harmonised Standards for the European AI Act, JRC Policy Brief, JRC 139430, 2024.
- Sommerer, L.M., Geospatial Predictive Policing – Research Outlook & A Call for Legal Debate, in *Neue Kriminalpolitik*, 29(2), 2017, 147-164.
- Sovrano, F., Sapienza, S., Palmirani, M., Vitali, F., Metrics, Explainability and the European AI Act Proposal, in *J*, 5(1), 2022, 126-138.
- Spiller, E., Il diritto di comprendere, il dovere di spiegare. Explainability e intelligenza artificiale costituzionalmente orientata, in *BioLaw Journal*, 2, 2021, 419-433.
- Sprenger, J., Brodowski, D., 'Predictive Policing', 'Predictive Justice', and the Use of 'Artificial Intelligence' in the Administration of Criminal Justice in Germany, *Association internationale de droit pénal*, 2023.
- Stanford University, AI Index Report 2024, 2024.
- Stark, L., Facial Recognition Is the Plutonium of AI, in *XRDS: Crossroads, The ACM Magazine for Students*, 25(3), 2019, 50-55.

BIBLIOGRAPHY

- Stefan, O., Avbelj, M., Eliantonio, M., Hartlapp, M., Korkea-Aho, E., Rubio, N., EU Soft Law in the EU Legal Order: A Literature Review, in King's College London Law School Research Paper, 9, 2019.
- Steiker, C.S., Foreword: The Limits of the Preventive State, in *The Journal of Criminal Law and Criminology*, 88(3), 1998, 771-808.
- Strikwerda, L., Predictive Policing: The Risks Associated with Risk Assessment, in *The Police Journal: Theory, Practice and Principles*, 94(3), 2021, 422-441.
- Strozzi, A. (Ed.), *Sicurezza nazionale: Poteri, conflitti, informazioni*, Luiss University Press, 2024.
- Taigman, Y., Yang, M., Ranzato, M., Wolf, L., DeepFace: Closing the Gap to Human-Level Performance in Face Verification, 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, 1701-1708.
- Tayebi, M.A., Glässer, U., *Social Network Analysis in Predictive Policing*, Springer, 2016.
- Taylor, L., Safety in Numbers? Group Privacy and Big Data Analytics in the Developing World, in L. Taylor, B. Van Der Sloot, L. Floridi (Eds.), *Group Privacy: The Challenges of New Data Technologies*, Springer, 2017, 13-36.
- Terrano, G., *La discrezionalità amministrativa nell'evoluzione della forma di Stato*, Doctoral Thesis, 2023.
- Thampi, A., *Interpretable AI. Building Explainable Machine Learning Systems*, Manning, 2022.
- Thornburg, R.H., Face Recognition Technology: The Potential Orwellian Implications and Constitutionality of Current Uses Under the Fourth Amendment, in *John Marshall Journal of Information Technology & Privacy Law*, 20(2), 2002, 321-346.
- Tipik Legal and Milieu Consulting, *Study on the Transposition of Directive (EU) 2016/680*, March 2023.
- Toledo, W.A., Murillo, A.M., The Obligation to Keep Records Of High-Risk Systems In The Artificial Intelligence Act, in L. Cotino Hueso, D.-U. Galetta (Eds.), *The European Union Artificial Intelligence Act. A Systematic Commentary*, CERIDAP, Editoriale Scientifica, 2025, 649-663.
- Torchia, L., La giustizia amministrativa digitale, in M. Ramajoli (Ed.), *Una giustizia amministrativa digitale?*, Il Mulino, 2023.

BIBLIOGRAPHY

- Torchia, L., *Lo Stato digitale: Una introduzione*, Il Mulino, 2025.
- Torchia, L., *Poteri pubblici e poteri privati nel mondo digitale*, Il Mulino, 2024.
- Torchia, L., *Pubblica Amministrazione e transizione digitale*, *Giornale di Diritto Amministrativo*, 6, 2024, 729-735.
- Tosoni, L., Bygrave, L.A., Article 3. Definitions, in E. Kosta, F. Boehm (Eds.), *The EU Law Enforcement Directive (LED): A Commentary*, Oxford University Press, 2024, 99-123.
- Townsley, M., Homel, R., Chaseling, J., *Infectious Burglaries: A Test of the Near Repeat Hypothesis*, in *British Journal of Criminology*, 43(3), 2003, 615-633.
- Townsley, M., Homel, R., Chaseling, J., *Repeat Burglary Victimization: Spatial and Temporal Patterns*, in *Australian and New Zealand Journal of Criminology*, 33(1), 2000, 37-63.
- Traverso, P., *Breve introduzione tecnica all'Intelligenza Artificiale*, in *DPCE online*, 1, 2022, 155-167.
- Trincado Castán, C., *The Legal Concept of Artificial Intelligence: The Debate Surrounding the Definition of AI System in the AI Act*, in *BioLaw Journal*, 1, 2024, 305-344.
- Tsaoussi, A., *Bounded Rationality*, in J. Backhaus (Ed.), *Encyclopedia of Law and Economics*, Springer, 2014.
- Turing, A.M., *Computing Machinery and Intelligence*, in *Mind*, 59(236), 1950, 433-460.
- Turing, A.M., *On Computable Numbers, with an Application to the Entscheidungsproblem*, in *Proceedings of the London Mathematical Society*, s2-42(1), 1937, 230-265.
- Turk, M., Pentland, A., *Eigenfaces for Recognition*, in *Journal of Cognitive Neuroscience*, 3(1), 1991, 71-86.
- Tyler, T.R., Fagan, J., *Legitimacy and Cooperation: Why Do People Help the Police Fight Crime in Their Communities?*, in *Ohio State Journal of Criminal Law*, 6, 2008, 231-275.
- Tyler, T.R., *Why People Obey The Law*, Yale University Press, 1990.
- Uchida, C.D., *A National Discussion on Predictive Policing: Defining Our Terms and Mapping Successful Implementation Strategies*, November 18-19, 2009.
- Uchida, C.D., *Predictive Policing*, in G. Bruinsma, D. Weisburd (Eds.), *Encyclopedia of Criminology and Criminal Justice*, Springer, 2014, 3862-3871.

BIBLIOGRAPHY

- Uchida, C.D., Swatt, M., Gamero, D., Lopez, J., Salazar, E., King, E., Maxey, R., Ong, N., Wagner, D., White, M.D., Los Angeles, California Smart Policing Initiative: Reducing Gun-Related Violence Through Operation LASER, 2012.
- UK Government, The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023.
- UNE-EN ISO/IEC 27001:2023 (ISO/IEC 27001:2022), Information Security Management System Requirements.
- UNESCO, Recommendation on the Ethics of Artificial Intelligence, 2021.
- UNICRI, Interpol, Responsible AI Innovation in Law Enforcement: AI Toolkit, February 2024.
- United Nations High Commissioner for Human Rights, Impact of New Technologies on the Promotion and Protection of Human Rights in the Context of Assemblies, Including Peaceful Protests, UN Doc. A/HRC/44/24, Report, 24 June 2020.
- UNODC, Handbook on Police Accountability, Oversight and Integrity, United Nations, July 2011.
- Ursi, R., La sicurezza nelle città: dallo spazio difendibile a tecnopolis, in *Munus*, 2, 2024, 1-31.
- Urwin, S., Algorithmic Forecasting of Offender Dangerousness for Police Custody Officers: An Assessment of Accuracy for the Durham Constabulary Model, Master Thesis in Applied Criminology and Police Management, University of Cambridge, 2016.
- Vaccaro, M., Waldo, J., The Effects of Mixing Machine Learning and Human Judgment, in *Communication of the ACM*, 62(11), 2019, 104-110.
- Vale, D., El Sharif, A., Ali, M., Explainable Artificial Intelligence (XAI) Post-Hoc Explainability Methods: Risks and Limitations in Non-Discrimination Law, in *AI & Ethics*, 2(1), 2022, 1-12.
- Vallor, S., Green, B., Raicu, I., Overview of Ethics in Tech Practice, Project Ethics in Technology Practice of Markkula Center for Applied Ethics, June 2018.
- Vallor, S., Green, B., Raicu, I., Technology and Engineering Practice: Ethical Lenses to Look Through, Project Ethics in Technology Practice of Markkula Center for Applied Ethics, July 2020.
- Van Bekkum, M., Borgesius, F.Z., Digital Welfare Fraud Detection and the Dutch SyRI Judgment, in *European Journal of Social Security*, 23(4), 2021, 323-340.

BIBLIOGRAPHY

- Van Brakel, R., De Hert, P., Policing, Surveillance and Law in a Pre-Crime Society: Understanding the Consequences of Technology Based Strategies, in *Cahiers Politiestudies*, 3(20), 2011, 163-192.
- Van Brakel, R., Pre-Emptive Big Data Surveillance and its (Dis)Empowering Consequences: The Case of Predictive Policing, in B. Van der Sloot et al. (Eds.), *Exploring the Boundaries of Big Data*, Amsterdam University Press, 2016, 117-141.
- Van der Hulst, R., Introduction to Social Network Analysis (SNA) as an Investigative Tool, in *Trends in Organized Crime*, 12, 2009, 101-121.
- Van Dijck, J., Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology, in *Surveillance & Society*, 12(2), 2014, 197-208.
- Van Loon, A., Toshkov, D., Adopting Open Source Software in Public Administration: The Importance of Boundary Spanners and Political Commitment, in *Government Information Quarterly*, 32(2), 2015, 207-215.
- Van Natta, M. et al., The Rise and Regulation of Thermal Facial Recognition Technology during the COVID-19 Pandemic, in *Journal of Law and the Biosciences*, 7(1), 2020, Article Isaa038.
- Varshney, K.R., *Trustworthy Machine Learning*, Independently Published, Chappaqua, NY, 2022.
- Veale, M., Binns, R., Edwards, L., Algorithms That Remember: Model Inversion Attacks and Data Protection Law, in *Philosophical Transactions of the Royal Society*, 376(2133), 2018, Article 20180083, 1-16.
- Veale, M., Zuiderveen Borgesius, F., Demystifying the Draft EU Artificial Intelligence Act, in *Computer Law Review International*, 22(4), 2021, 97-112.
- Venema, R., How to Govern Visibility?: Legitimizations and Contestations of Visual Data Practices after the 2017 G20 Summit in Hamburg, in *Surveillance & Society*, 18(4), 2020, 522-539.
- Verdecchia, R., Sallou, J., Cruz, L., A Systematic Review of Green AI, in *WIREs Data Mining and Knowledge Discovery*, 13(4), 2023, Article e1507, 1-43.
- Vogiatzoglou, P., Fantin, S., National and Public Security Within and Beyond the Police Directive, in A. Vedder, J. Schroers, C. Ducuing, P. Valcke (Eds.), *Security and Law: Legal and Ethical Aspects of Public Security, Cyber Security and Critical Infrastructure Security*, Intersentia, 2019, 123-156.

BIBLIOGRAPHY

- Vogiatzoglou, P., Marquenie, T., Assessment of the Implementation of the Law Enforcement Directive, Study Requested by the European Parliament's Committee on Civil Liberties, Justice and Home Affairs, December 2022.
- Vogiatzoglou, P., Quezada Tavárez, K., Fantin, S., Dewitte, P., From Theory to Practice: Exercising the Right of Access under the Law Enforcement and PNR Directives, in *Journal of Intellectual Property, Information Technology and E-Commerce Law*, 11(3), 2020, 274-302.
- Vogiatzoglou, P., The Era of Pre-Crime: How Mass Data Surveillance and Predictive Policing Intersect and Interfere with Privacy, Data Protection and Due Process Rights in the EU, Doctoral Thesis, KU Leuven Faculty of Law and Criminology, 2023.
- Volpato, A., Eliantonio, M., The Contradictory Approach of the CJEU to the Judicial Review of Standards: A Love-Hate Relationship?, in M. Eliantonio, C. Cauffman (Eds.), *The Legitimacy of Standardisation as a Regulatory Technique: A Cross-disciplinary and Multi-level Analysis*, Edward Elgar Publishing, 2020, 91-109.
- Volpato, A., The Legal Effects of Harmonised Standards in EU Law: From Hard to Soft Law, and Back?, in *Review of European Administrative Law*, 16(2), 2023, 5-35.
- Volpato, A., The Publicity of EU Law and the Privatization of EU Digital Regulation, in *Maastricht Journal of European and Comparative Law*, 31(3), 2024, 319-332.
- Von Neumann, J., Morgenstern, O., *The Theory of Games and Economic Behavior*, Princeton University Press, 2007.
- Wachter, S., Mittelstadt, B., Russell, C., Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR, in *Harvard Journal of Law & Technology*, 31(2), 2018, 841-887.
- Wang, M., Deng, W., Deep Face Recognition: A Survey, arXiv:1804.06655, 2020.
- Wang, R.Y., Strong, D.M., Beyond Accuracy: What Data Quality Means to Data Consumers, in *Journal of Management Information Systems*, 12(4), 1996, 5-33.
- Wang, Y., Kosinski, M., Deep Neural Networks Are More Accurate than Humans at Detecting Sexual Orientation from Facial Images, in *Journal of Personality and Social Psychology*, 114(2), 2018, 246-257.
- Waxman, M.C., Police and National Security: American Local Law Enforcement and Counter-Terrorism after 9/11, in *Journal of National Security Law and Policy*, 3, 2009, 377-407.

BIBLIOGRAPHY

- Weatherill, S., Harmonization in the EU Internal Market, in *Oxford Encyclopedia of EU Law*, Oxford University Press, 2024, 1-28.
- Weatherill, S., The Fundamental Question of Minimum or Maximum Harmonisation, in S. Garben, I. Govaere (Eds.), *The Internal Market 2.0*, Oxford University Press, 2020, 261-282.
- Weatherill, S., *The Internal Market as a Legal Concept*, Oxford University Press, 2016.
- Weatherill, S., The Limits of Legislative Harmonization Ten Years after Tobacco Advertising: How the Court's Case Law Has Become a “Drafting Guide”, in *German Law Journal*, 12(3), 2011, 827-864.
- Weisburd, D., Mastrofski, S., Willis, J.J., Greenspan, R., Changing Everything So That Everything Can Remain the Same: Compstat and American Policing, in D. Weisburd, A.A. Braga (Eds.), *Police Innovation: Contrasting Perspectives*, Cambridge University Press, 2001, 284-301.
- Weisburd, D., The Law of Crime Concentration and the Criminology of Place, in *Criminology*, 53(2), 2015, 133-157.
- Weizenbaum, J., ELIZA—A Computer Program for the Study of Natural Language Communication between Man and Machine, in *Communications of the ACM*, 9(1), 1966, 36-45.
- Welsh, B.C., Farrington, D.P., Crime Prevention and Public Policy, in D.P. Farrington, B.C. Welsh (Eds.), *The Oxford Handbook of Crime Prevention*, Oxford University Press, 2012, 3-19.
- Wendehorst, C., Duller, Y., Biometric Recognition and Behavioural Detection: Assessing the Ethical Aspects of Biometric Recognition and Behavioural Detection Techniques with a Focus on Their Current and Future Use in Public Spaces, Study commissioned by the European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs, August 2021.
- Westin, A., Privacy and Freedom, in *Administrative Law Review*, 22(1), 1969, 101-106.
- Wiedemann, K., Profiling and (Automated) Decision-Making under the GDPR: A Two-Step Approach, in *Computer Law & Security Review*, 45, 2022, Article 105670, 1-14.
- Wiener, N., *Cybernetics: Or Control and Communication in the Animal and the Machine*, MIT Press, 1961.
- Wiener, N., Some Moral and Technical Consequences of Automation, in *Science*, 131(3410), 1960, 1355-1358.

BIBLIOGRAPHY

- Williams, P., Kind, E., Data-Driven Policing: The Hardwiring of Discriminatory Policing Practices across Europe, Report, European Network Against Racism, 2019.
- Willis, J.J., Mastrofski, S.D., Weisburd, D., Compstat in Practice: An In-Depth Analysis of Three Cities, Police Foundation, 2003.
- Winston, P.H., Artificial Intelligence, Addison-Wesley, 1992.
- Witten, I.H., Frank, E., Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2011.
- Wu, X., Zhang, X., Automated Inference on Criminality Using Face Images, arXiv:1611.04135v1, 26 May 2017.
- Xenidis, R., Senden, L., EU Non-Discrimination Law in the Era of Artificial Intelligence: Mapping the Challenges of Algorithmic Discrimination, in U. Bernitz et al. (Eds.), General Principles of EU Law and the EU Digital Order, Kluwer Law International, 2020, 151-182.
- Xenidis, R., Tuning EU Equality Law to Algorithmic Discrimination: Three Pathways to Resilience, in Maastricht Journal of European and Comparative Law, 27(6), 2020, 736-758.
- Yi, D., Lei, Z., Liao, S., Li, S.Z., Learning Face Representation from Scratch, arXiv:1411.7923, 2014.
- Yule, G.U., On the Methods of Measuring Association Between Two Attributes, in Journal of the Royal Statistical Society, 75(6), 1912, 579-652.
- Zaki, M.J., Wagner, M., Data Mining and Machine Learning: Fundamental Concepts and Algorithms, Cambridge University Press, 2020.
- Zalnieriute, M., Facial Recognition Surveillance and Public Space: Protecting Protest Movements, in International Review of Law, Computers & Technology, 39(1), 2025, 116–135.
- Zamfir, I. (European Parliamentary Research Service), Definitions of Rape in the Legislation of EU Member States, May 2025, PE 757.618.
- Zanichelli, M., Ecosistemi, opacità, autonomia: le sfide dell'intelligenza artificiale in alcune proposte recenti della Commissione europea, in A. D'Aloia (Ed.), Intelligenza artificiale e diritto. Come regolare un mondo nuovo, FrancoAngeli, 2020.
- Zansberg, S.D., Campos, P., Sunshine on the Thin Blue Line: Public Access to Police Internal Affairs Files, in Communications Lawyer, 2004, 34-39.

BIBLIOGRAPHY

- Završnik, A., Blurring the Line between Law Enforcement and Intelligence: Sharpening the Gaze of Surveillance?, in *Journal of Contemporary European Research*, 9(1), 2013, 181-202.
- Zedner, L., Pre-Crime and Post Criminology?, in *Theoretical Criminology*, 11(2), 2007, 261-281.
- Zeng, J., Ustun, B., Rudin, C., Interpretable Classification Models for Recidivism Prediction, in *Journal of the Royal Statistical Society*, 180(3), 2017, 689-722.
- Zhang, Hong-Bo, Zhang, Yi-Xiang, Zhong, Bineng, Lei, Qing, Yang, Lijie, Du, Ji-Xiang, Chen, Duan-Sheng, A Comprehensive Survey of Vision-Based Human Action Recognition Methods, in *Sensors*, 19(5), 2019, 1005.
- Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A., Face Recognition: A Literature Survey, in *ACM Computing Surveys*, 35(4), 2003, 399-458.
- Zuboff, S., *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, PublicAffairs, 2019.
- Zuccon, G., Scells, H., Zhuang, S., Beyond CO2 Emissions: The Overlooked Impact of Water Consumption of Information Retrieval Models, in *Proceedings of the 2023 ACM SIGIR International Conference on Theory of Information Retrieval*, 2023, 283-289.
- Zuiderveen Borgesius, F., *Discrimination, Artificial Intelligence and Algorithmic Decision-Making*, Council of Europe's European Commission Against Racism and Intolerance (ECRI), 2018.

Laws and Regulations

Supranational

Charter of Fundamental Rights of the European Union, OJ C 202, 7 June 2016.

Consolidated version of the Treaty on the Functioning of the European Union, OJ C 326, 26 October 2012.

Consolidated version of the Treaty on the Functioning of the European Union – Protocol (No 2) on the application of the principles of subsidiarity and proportionality, Art. 5, OJ C 115, 9 May 2008.

Consolidated version of the Treaty on the European Union, OJ C 326, 26 October 2012.

Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data by Competent Authorities for the Purposes of the Prevention, Investigation, Detection or Prosecution of Criminal Offences or the Execution of Criminal Penalties (Law Enforcement Directive - LED).

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation - GDPR).

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/182 (Artificial Intelligence Act – AI Act). Official Journal of the European Union, L series, 1-144, 12 July 2024.

Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European Standardisation.

Proposal for a Regulation amending Regulations (EU) 2024/1689 and (EU) 2018/1139 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI), COM(2025) 836 final, 19 November 2025.

Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), COM(2021) 206 final, 2021/0106(COD), 21 April 2021.

BIBLIOGRAPHY

Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).

Proposal for a Regulation on Artificial Intelligence - Council of the European Union, General Approach, Interinstitutional File: 2021/0106(COD), 25 November 2022.

National

City and County of San Francisco, Board of Supervisors Approval of Surveillance Technology Policy, San Francisco Administrative Code Section 19B.2(d), 2019.

Décret n° 2021-148 du 11 février 2021 portant modalités de mise en œuvre par la direction générale des finances publiques et la direction générale des douanes et droits indirects de traitements informatisés et automatisés permettant la collecte et l'exploitation de données rendues publiques sur les sites internet des opérateurs de plateforme en ligne.

Décret n° 2023-828 du 28 août 2023 relatif aux modalités de mise en œuvre des traitements algorithmiques sur les images collectées au moyen de systèmes de vidéoprotection et de caméras installées sur des aéronefs.

Décret n° 2023-939 du 11 octobre 2023 relatif aux modalités de pilotage et d'évaluation de l'expérimentation de traitements algorithmiques d'images légalement collectées au moyen de systèmes de vidéoprotection et de caméras installées sur des aéronefs.

Décret n° 2024-1274 du 31 décembre 2024 modifiant le décret n° 2021-148 du 11 février 2021 portant modalités de mise en œuvre par la direction générale des finances publiques et la direction générale des douanes et droits indirects de traitements informatisés et automatisés permettant la collecte et l'exploitation de données rendues publiques sur les sites internet des opérateurs de plateforme en ligne.

Decreto 10 maggio 1994, n. 415. Regolamento per la disciplina delle categorie di documenti sottratti al diritto di accesso ai documenti amministrativi, in attuazione dell'art. 24, comma 4, della legge 7 agosto 1990, n. 241, recante nuove norme in materia di procedimento amministrativo e di diritto di accesso ai documenti amministrativi.

BIBLIOGRAPHY

- Decreto del Ministro dell'Interno, 24 May 2017. Individuazione dei trattamenti di dati personali effettuati dal Centro elaborazione dati del Dipartimento della pubblica sicurezza o da Forze di polizia sui dati destinati a confluirci, ovvero da organi di pubblica sicurezza o altri soggetti pubblici nell'esercizio delle attribuzioni conferite da disposizioni di legge o di regolamento, effettuati con strumenti elettronici e i relativi titolari, in attuazione dell'articolo 53, comma 3, del decreto legislativo 30 giugno 2003, n. 196. (17A04275)
- Decreto del Presidente della Repubblica 15 gennaio 2018, n. 15. Regolamento a norma dell'articolo 57 del decreto legislativo 30 giugno 2003, n. 196, recante l'individuazione delle modalità di attuazione dei principi del Codice in materia di protezione dei dati personali relativamente al trattamento dei dati effettuato, per le finalità di polizia, da organi, uffici e comandi di polizia. (18G00040) D.P.R. 27 giugno 1992, n. 352.
- Decreto Legislativo 18 maggio 2018, n. 51, Attuazione della direttiva (UE) 2016/680 relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali (Transposition of the LED).
- Decreto Legislativo 31 marzo 2023, n. 36, Codice dei contratti pubblici (Code of public procurement).
- Hessisches Gesetz über die öffentliche Sicherheit und Ordnung (HSOG) in der Fassung der Bekanntmachung vom 14. January 2005.
- Legge 23 settembre 2025, n. 132, Disposizioni e deleghe al Governo in materia di intelligenza artificiale.
- Loi n° 2019-1479 du 28 décembre 2019 de finances pour 2020.
- Loi n° 2023-1322 du 29 décembre 2023 de finances pour 2024.
- Loi n° 2023-380 du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024 et portant diverses autres dispositions.
- State of Colorado, S113, 2022.
- State of Maryland, SB182, 2024.
- State of Washington, S6280, 2020.
- U.S. Congress, Senate Bill 2551, 117th Congress.

BIBLIOGRAPHY

Case Law

Supranational Courts

CJEU, Grand Chamber, Judgement of 21 December 2016, *Tele2 Sverige AB v. Post-och telestyrelsen*; *Secretary of State for the Home Department v. Watson and Others*, Joined Cases C-203/15 and C-698/15, ECLI:EU:C:2016:970.

CJEU, Grand Chamber, Judgement of 21 June 2022, *Ligue des Droits Humains v. Conseil des ministers*, Case C-817/19, ECLI:EU:C:2022:491.

CJEU, Grand Chamber, Judgement of 8 April 2014, *Digital Rights Ireland Ltd v. Minister for Communications*, Joined Cases C-293/12 and C-594/12, ECLI:EU:C:2014:238.

CJEU, Judgment of 1 August 2022, *OT v. Vyriausioji tarnybinės etikos komisija*, Case C-184/20, ECLI:EU:C:2022:601.

CJEU, Judgment of 13 June 2018, *Commission v Poland*, Case C-530/16, ECLI:EU:C:2018:430.

CJEU, Judgment of 16 July 2015, *'CHEZ Razpredelenie Bulgaria' AD v Komisija za zashtita ot diskriminatsia*, Case C-83/14, ECLI:EU:C:2015:480.

CJEU, Judgment of 17 July 2008, *S. Coleman v Attridge Law and Steve Law*, Case C-303/06, ECLI:EU:C:2008:415.

CJEU, Judgment of 19 October 2016, *Patrick Breyer v Bundesrepublik Deutschland*, Case C-582/14, ECLI:EU:C:2016:779.

CJEU, Judgment of 2 April 2020, *Commission v Poland, Hungary and the Czech Republic*, Joined Cases C-715/17, C-718/17 and C-719/17, ECLI:EU:C:2020:257.

CJEU, Judgment of 2 October 2018, *Ministerio Fiscal*, Case C-207/16, ECLI:EU:C:2018:788.

CJEU, Judgment of 20 December 2017, *Peter Nowak v Data Protection Commissioner*, Case C-434/16, ECLI:EU:C:2017:994.

CJEU, Judgment of 21 June 2022, *Ligue des droits humains*, Case C-817/19, ECLI:EU:C:2022:491.

CJEU, Judgment of 22 June 2021, *Latvijas Republikas Saeima (Penalty points)*, Case C-439/19, ECLI:EU:C:2021:504.

BIBLIOGRAPHY

- CJEU, Judgment of 22 May 2012, P.I. v Oberbürgermeisterin der Stadt Remscheid, Case C-348/09, ECLI:EU:C:2012:300.
- CJEU, Judgment of 23 November 2010, Land Baden-Württemberg v Panagiotis Tsakouridis, Case C-145/09, ECLI:EU:C:2010:708.
- CJEU, Judgment of 24 February 2022, SIA SS v Valsts ieņēmumu dienests, Case C-175/20, ECLI:EU:C:2022:124.
- CJEU, Judgment of 24 June 2015, H.T. v Land Baden-Württemberg, Case C-373/13, ECLI:EU:C:2015:413.
- CJEU, Judgment of 24 November 2016, David L. Parris v. Trinity College Dublin and Others, Case C-443/15, ECLI:EU:C:2016:897.
- CJEU, Judgment of 26 February 2013, Åklagaren v Hans Åkerberg Fransson, Case C-617/10, ECLI:EU:C:2013:105.
- CJEU, Judgment of 27 October 2016, James Elliott Construction Limited v. Irish Asphalt Limited, Case C-613/14, ECLI:EU:C:2016:821.
- CJEU, Judgment of 4 June 2013, ZZ v Secretary of State for the Home Department, Case C-300/11, ECLI:EU:C:2013:363.
- CJEU, Judgment of 4 September 2025, European Data Protection Supervisor v Single Resolution Board, Case C-413/23, ECLI:EU:C:2025:645.
- CJEU, Judgment of 5 April 2022, G.D. v The Commissioner of the Garda Síochána and Others, Case C-140/20, ECLI:EU:C:2022:258.
- CJEU, Judgment of 5 October 2000, Federal Republic of Germany v. European Parliament and Council of the European Union, Case C-376/98, ECLI:EU:C:2000:544.
- CJEU, Judgment of 6 October 2015, Maximillian Schrems v Data Protection Commissioner, Case C-362/14, ECLI:EU:C:2015:650.
- CJEU, Judgment of 6 October 2020, La Quadrature du Net and Others v Premier ministre and Others, Joined Cases C-511/18, C-512/18, C-520/18, ECLI:EU:C:2020:791.
- CJEU, Judgment of 6 October 2020, Privacy International v Secretary of State for Foreign and Commonwealth Affairs and Others, Case C-623/17, ECLI:EU:C:2020:790.

BIBLIOGRAPHY

- CJEU, Judgment of 7 December 2023, *OQ v Land Hessen*, Case C-634/21, ECLI:EU:C:2023:957.
- CJEU, Judgment of 7 May 2009, *College van burgemeester en wethouders van Rotterdam v M.E.E. Rijkeboer*, Case C-553/07, ECLI:EU:C:2009:293.
- CJEU, Judgment of 8 December 2021, *Inspektor V Inspektorata Kam Visshia Sadeben Savet*, Case C-180/21, ECLI:EU:C:2022:967.
- ECtHR Grand Chamber, Judgement of 4 December 2008, *S. and Marper v. United Kingdom*, Application Nos. 30562/04 and 30566/04.
- ECtHR, Grand Chamber, Judgment of 21 October 2013, *Janowiec and Others v. Russia*, Application nos. 55508/07 and 29520/09.
- ECtHR, Judgement of 24 January 2019, *Catt v. United Kingdom*, Application No. 43514/15.
- ECtHR, Judgment of 13 September 2018, *Big Brother Watch and Others v. the United Kingdom*, Application nos. 58170/13, 62322/14 and 24960/15.
- ECtHR, Judgment of 4 July 2023, *Glukhin v. Russia*, Application no. 11519/20.
- ECtHR, Judgment of 6 September 1978, *Klass and Others v. Germany*, Application no. 5029/71.
- ECtHR, Plenary, Judgment of 8 June 1976, *Engel and Others v. the Netherlands*, Application nos. 5100/71, 5101/71, 5102/71, 5354/72, 5370/72.
- ECtHR, Section II, Judgment of 4 March 2014, *Grande Stevens v. Italy*, Application no. 18640/10.
- ### National Courts
- Bundesverfassungsgericht, I Senate (President Stephan Harbarth), Judgment of 16 February 2023, 1 BvR 1547/19, 1 BvR 2634/20.
- Conseil constitutionnel, Decision no. 2018-765 DC, 12 June 2018.
- Conseil constitutionnel, Decision no. 2023-850 DC, 17 May 2023.
- Consent Decree, *United States v. Police Department of Baltimore City*, No. 1:17-cv-00099 JKB (D. Md. Jan. 12, 2017).
- Consiglio di Stato, Section IV, Judgment of 4 June 2025, No. 4857.
- Consiglio di Stato, Section VI, Judgment of 13 December 2019, Nos. 8472 and 8473.
- Consiglio di Stato, Section VI, Judgment of 4 February 2020, No. 881.

BIBLIOGRAPHY

Consiglio di Stato, Section VI, Judgment of 8 April 2019, No. 2270.

Corte Costituzionale, Judgment No. 129/2009.

Corte Costituzionale, Judgment No. 13/1994.

Corte Costituzionale, Judgment No. 218/1988.

Corte Costituzionale, Judgment No. 290/2001.

Floyd, et al. v. City of New York, et al., 959 F. Supp. 2d 540 (S.D.N.Y. 2013).

Hague District Court, NJCM et al. v The Dutch State (SyRI), 5 February 2020.

Supreme Court of Wisconsin, State of Wisconsin v. Eric L. Loomis, Case no. 2015AP157-CR, 2016.

TAR Lazio-Roma, Judgment of 30 June 2021, No. 7769.

TAR Lazio-Roma, Section III-bis, Judgment of 19 April 2019, No. 5139.

TAR Lazio-Roma, Section III-bis, Judgment of 30 June 2020, No. 7370.

Tribunal Administratif de Marseille, Judgment of 27 February 2020, N° 1901249.

Tribunal Supremo, Sala de lo Contencioso-Administrativo, Sección Tercera, Judgment No. 1119/2025, 11 September 2025.

UK Court of Appeal, Civil Division, R (Bridges) v. Chief Constable of South Wales Police, Case C1/2019/2670, 11 August 2020.

UK High Court of Justice, Queen's Bench Division, Divisional Court Cardiff, Judgement of 4 September 2019, R (Bridges) v Chief Constable of South Wales Police and Others, EWHC 2341.

La borsa di dottorato è stata cofinanziata con risorse dell'Unione europea-*NextGeneration EU*
Piano Nazionale di Ripresa e Resilienza (PNRR)