UNIVERSITÀ DEGLI STUDI DI TRENTO
FACOLTÀ DI GIURISPRUDENZA
Dottorato in Studi Giuridici
Comparati ed Europei

Corso di Dottorato in Studi Giuridici Comparati ed Europei

XXXV ciclo

Tesi di Dottorato

L'intelligenza artificiale come nuova frontiera dei diritti fondamentali

Relatori
Prof. Carlo Casonato
Dott. Paolo Traverso

Dottorando
Luca Rinaldi

anno accademico 2022-2023

UNIVERSITÀ DEGLI STUDI DI TRENTO
FACOLTÀ DI GIURISPRUDENZA
Dottorato in Studi Giuridici
Comparati ed Europei

candidato: Luca Rinaldi

L'INTELLIGENZA ARTIFICIALE COME NUOVA FRONTIERA DEI DIRITTI FONDAMENTALI

relatori: Prof. Carlo Casonato
Dott. Paolo Traverso

Anno Accademico 2022-2023

Curriculum di Diritto amministrativo, costituzionale e internazionale

XXXV ciclo

Esame finale: 16 marzo 2023

Commissione esaminatrice: prof. Antonio D'Aloia, Università di Parma
prof. Federico Gustavo Pizzetti, Università di Milano
prof.ssa Elettra Stradella, Università di Pisa

Marzo 2023

La scrittura di questa tesi è stata il principale impegno degli ultimi tre anni della mia vita. Un periodo che ricorderò sempre con piacere, ricco di stimoli e opportunità: ho scoperto idee, persone e luoghi fantastici, dei quali sarò sempre grato alla vita. Un dottorato di ricerca, però, richiede anche una considerevole dose di impegno e volontà, in cui trovare la forza di proseguire quando la curiosità sembra non bastare più. Per me, almeno, è stato così. Le persone che abbiamo accanto, in quei momenti, sono decisive, ed è in larga parte da loro che dipende la buona riuscita del lavoro. Senza gli uomini e le donne cui devo questi ringraziamenti, queste pagine non esisterebbero.

Devo ringraziare innanzitutto i miei tutor, Carlo Casonato e Paolo Traverso, per aver creduto in me e avermi offerto, col coraggio delle loro scelte, la possibilità di lavorare ai confini tra discipline distinte. In particolare, sono grato al primo per i preziosi consigli e il continuo supporto durante la scrittura, e al secondo per l'opportunità di frequentare il mondo della Fondazione Bruno Kessler e conoscere gli eccellenti professionisti che lavorano al suo interno.
Devo un grande grazie, poi, all'intero gruppo di ricerca BioDiritto (Carla, Cinzia, Elisabetta, Giulia, Lucia, Marta I, Marta II, Sergio e Simone) per avermi fatto sentire a casa fin dal primo giorno, per i continui stimoli e per le tante occasioni di lavoro di squadra, senza mai perdere il sorriso. Marta, in particolare, è stata una guida preziosa e un fedele alleato in cui specchiarmi per consigli e conforto nelle difficoltà. Devo molto anche a Andrea, Lorenzo, Monica e al resto del team di Trentino Salute 4.0, cui va la mia gratitudine per l'accoglienza che mi hanno riservato in un ambiente per me del tutto nuovo e per le innumerevoli occasioni di contaminazione interdisciplinare che mi hanno fornito. Sarò sempre debitore, per la stessa ragione, anche a Paolo, Giorgia e Federico, come me giuristi legati alla Fondazione, una riserva inesauribile di idee, stimoli e consigli.
In più, mi sono stati di grande aiuto i ricercatori del Max Planck Institute for Social Law and Social Policy di Monaco di Baviera e del Center for Biomedical Innovation Law dell'Università di Copenhagen, che hanno reso i miei periodi di ricerca all'estero delle preziose occasioni di confronto con giuristi di altri ordinamenti. Per tali opportunità devo ringraziare, innanzitutto, i professori Ulrich Becker, Timo Minnsen e Marcelo Corrales Compagnucci, che dirigono quei centri e mi hanno magnificamente accolto.

Poi, la mia famiglia - che mi supporta e sopporta, silenziosamente, da quando sono nato - e i miei amici, tutti. Angela, Anna, Claudia, Claudio, Davide, Francesco, Francesca, Giorgio, Giovanni, Giulio, Irene, Luca, Martina, Roberto e gli altri con cui ho diviso il cammino fin da bambino; Anna, Alberto, Antonella, Chiara, Christian, Clara, Federico, Giorgia, Giulia, Michele, Orlane, Sara, Simona, che sono venuti dopo, ma mi sembra ci siano da sempre; i ragazzi di CNR; Mario, Alessandro e tutta la Dinamo Kave; i compagni di squadra diventati, grazie al calcio, compagni di strada; Berdien, Federico, Giovanni, Marta e Matteo, per l'amicizia sincera in un mondo in cui è rara; chi ho perso di vista, ma il cui ricordo renderà sempre speciali questi anni.

Infine, Elena. Lei sa perché.

INDICE

ABSTRACT

Il lavoro indaga le conseguenze dell'avvento dell'intelligenza artificiale sulla garanzia e l'effettività dei diritti fondamentali riconosciuti nella tradizione costituzionale italiana e in quelle ad essa accostabili. La prima parte dell'opera contestualizza l'argomento e si occupa, in larga misura, di discipline diverse dal diritto. Innanzitutto, è offerta una ricostruzione sintetica delle caratteristiche delle tecnologie ascritte alla famiglia dell'intelligenza artificiale, delle possibili definizioni di quest'ultima e dei momenti essenziali del suo sviluppo. In seguito, il testo ripercorre, per brevi cenni, il dibattito filosofico e sociologico sul ruolo della tecnica nella dimensione umana ed elenca gli interrogativi più discussi in materia di etica della tecnologia e rapporto uomo-macchina. Gli ultimi paragrafi inquadrano le principali ipotesi di regolazione del fenomeno allo stato dell'arte.

La seconda e la terza parte, invece, sono dedicate all'analisi giuridica. La seconda parte investiga l'impatto delle trasformazioni connesse all'avvento dell'intelligenza artificiale su alcune posizioni giuridiche fondamentali: i diritti a protezione della sfera dell'identità, la libertà di manifestazione del pensiero e il principio di eguaglianza. La scelta è ricaduta su queste istanze per l'intensità con cui l'intelligenza artificiale mette in discussione il quadro giuridico posto a loro protezione, rivelandone l'insufficienza e la necessità di rinnovamento. Il lavoro si sofferma sui principali problemi sollevati dai sistemi avanzati e, nell'esaminare le possibili soluzioni *de iure condendo*, prende in considerazione le ipotesi allo studio in diversi ordinamenti, compreso quello eurounitario.

La terza e ultima parte sviluppa una precisa posizione teorica: la rivoluzione causata dall'intelligenza artificiale ha effetti così dirompenti da rendere necessario teorizzare – e tutelare al massimo livello – situazioni giuridiche del tutto nuove, difficilmente riconducibili all'alveo applicativo di diritti esistenti. Il tema affrontato, dunque, è quello - risalente - della configurabilità di *nuovi diritti*, la cui tutela sia resa necessaria dall'evoluzione sociale, culturale e tecnologica, e dai pericoli per la sfera della personalità individuale che questa porta con sé. Le nuove posizioni giuridiche analizzate sono il diritto di conoscere la natura artificiale di un interlocutore, il diritto a una spiegazione dei risultati delle tecnologie avanzate e il diritto a una soglia minima di controllo umano su queste ultime. La scelta è caduta su di esse perché sembrano le più adatte a indirizzare lo sviluppo tecnologico verso un'intelligenza artificiale genuinamente *human-centred*, eliminando il rischio di ridurre l'uomo a oggetto di macchine che non comprende, non controlla o nemmeno percepisce. Al fine di dare concretezza alla riflessione, le ultime pagine dello studio calano i tre nuovi diritti in altrettanti scenari applicativi, scelti per la particolare rilevanza dei diritti individuali in gioco: la pubblica amministrazione, il sistema giustizia e l'attività medico-sanitaria.

Introduzione

Risulta difficile, oggi, trascorrere un'intera giornata senza sentir parlare di "intelligenza artificiale". Viviamo nel mezzo di un picco di attenzione per questa famiglia di tecnologie, originato dai successi raggiunti dal *deep learning* nella seconda metà dello scorso decennio. L'interesse proviene, innanzitutto, dai media, che a volte abusano del concetto, impiegandolo per tecnologie che intelligenti non sono. Anche la ricerca scientifica, però, dimostra particolare attenzione per l'argomento, non solo in campo tecnologico e ingegneristico. L'intelligenza artificiale, infatti, è da qualche tempo uno dei temi principali della riflessione antropologica, filosofica, psicologica e giuridica, al fine di indagarne il possibile impatto sull'uomo e sulla società.

Questo studio fa parte di questo filone di ricerca e consiste in un'indagine d'ampio respiro delle conseguenze dell'avvento dell'intelligenza artificiale sui diritti fondamentali. Si tratta, dunque, di uno studio riguardante una delle categorie del diritto più risalenti e consolidate. Ciò nonostante, la prima, basilare necessità, al momento di affrontare, da qualunque punto di vista, il tema dell'intelligenza artificiale, è, banalmente, *capire ciò di cui si parla*. Questo lavoro non fa eccezione, e, dunque, è aperto da una ricostruzione sintetica, ma rigorosa, delle possibili definizioni, della storia e delle principali caratteristiche tecniche dell'intelligenza artificiale. Le ragioni della scelta stanno nella necessità di delimitare i confini del concetto, resi incerti dal menzionato clamore mediatico che lo circonda, e approfondire le qualità delle tecnologie il cui impatto è indagato, dal punto di vista del diritto, nelle parti successive.

Impadronirsi delle nozioni tecniche necessarie per parlare, da un punto di vista che tecnico non è, di intelligenza artificiale è un'operazione lunga, faticosa e complessa. A renderla possibile è il dialogo con gli specialisti della disciplina, e in particolare con quelli tra di essi più attenti ai possibili risvolti sociali dell'innovazione tecnologica, e dunque interessati, a loro volta, al confronto con esperti di altri ambiti. Un ruolo decisivo, in questo processo di contaminazione, è rivestito da Università e centri di ricerca pubblici e privati, nei quali il numero di progetti e iniziative interdisciplinari continua a crescere, in modo lento ma costante. Un altro impulso fondamentale viene dal mondo istituzionale, in seno al quale è sempre più frequente la formazione di organi in cui specialisti di diverse discipline sono posti l'uno accanto all'altro, al fine specifico di indagare le possibili conseguenze etiche, giuridiche, sociali dell'avvento dell'intelligenza artificiale. Sullo scenario europeo, come si vedrà, uno dei documenti principali attorno a cui si svolge il dibattito sulla regolazione di questa nuova famiglia di tecnologie, le *Linee guida etiche per un'intelligenza*

artificiale affidabile, proviene proprio da un gruppo di esperti di questo tipo, appositamente costituito dalla Commissione.

Le conoscenze interdisciplinari di cui impadronirsi riguardano anche la discussione che sociologi, psicologi e filosofi portano avanti, in parallelo ai giuristi, sugli effetti dell'innovazione tecnologica, affrontando, a loro volta, le difficoltà generate dalla necessità di interfacciarsi con discipline diverse dalla propria. Un'indagine che voglia dirsi completa non può ignorare tali ricerche, e, una volta esaurito l'inquadramento tecnico appena descritto, questo lavoro dà conto, per brevi cenni essenziali, dei loro risultati e delle loro prospettive principali.

Preme ribadire, tuttavia, che la ricerca qui introdotta è una ricerca d'ambito giuridico, dedicata a una categoria ben precisa: i diritti fondamentali. L'approfondimento di altri campi della conoscenza è stato condotto col massimo rigore, e talvolta occupa, come si vedrà, diverse pagine, ma è sempre strumentale all'analisi in punto di diritto che anima l'intero lavoro, al fine di dotarla di senso, determinandone l'oggetto.

Del resto, proprio la volontà di evitare il pericolo di cadere in pericolose forme di dilettantismo interdisciplinare è stata tra le principali ragioni che hanno spinto a scegliere, come tema di lavoro, i diritti fondamentali, uno dei concetti giuridici essenziali più risalenti. Infatti, solo l'analisi di una categoria giuridica tanto basilare poteva mettere al riparo dal rischio che la ricostruzione del funzionamento di tecnologie tanto innovative e complesse finisse per risultare la vera protagonista dell'opera, relegando ai margini le considerazioni in diritto. La seconda ragione che ha portato a scegliere i diritti fondamentali come chiave di lettura dello studio, invece, si radica in una constatazione di fatto: le trasformazioni connesse all'avvento dell'intelligenza artificiale sono di tale intensità da mettere in discussione lo stesso rapporto tra uomo e natura e tra persona e società, chiamando in causa le categorie giuridiche di livello più alto. Al fine di sottolineare con forza questa circostanza, la ricerca adotta un approccio di ampio respiro, indagando l'impatto dell'intelligenza artificiale su diversi diritti fondamentali, fino a chiedersi se il catalogo di diritti patrimonio comune degli odierni stati costituzionali di diritto – al netto delle differenze, anche di rilievo, esistenti tra distinti ordinamenti – sia sufficiente o necessiti, oggi, di un integrazione di fronte alle sfide poste dall'intelligenza artificiale.

Così, dopo la prima parte, riservata, come già detto, all'esame degli aspetti tecnici e della storia dell'intelligenza artificiale, al dibattito filosofico e sociologico sul tema, e a un inquadramento delle principali ipotesi di regolazione del fenomeno sullo scenario globale, il corpo centrale del lavoro è dedicato ai diritti fondamentali. La seconda parte, in particolare, si concentra sulle sfide che questa nuova famiglia di tecnologie pone ad alcuni dei presidi giuridici individuali più consolidati nella nostra tradizione giuridica e in quelle ad essa accostabili. L'analisi verte, in primo luogo, sulle

conseguenze causate dall'avvento delle tecnologie intelligenti per i diritti attinenti alla sfera dell'identità personale. L'evoluzione generata, in materia, dalla digitalizzazione della società è ben nota, ed è stata commentata dalla migliore dottrina fin dagli anni '70 – il rimando è, prima di tutto, agli studi di Stefano Rodotà. Basti pensare alle tensioni a cui è stata sottoposta la tutela della riservatezza, cui è stato necessario far fronte con l'elaborazione teorica, e il riconoscimento in via legislativa e giurisprudenziale, di diritti fondamentali squisitamente "digitali", come il controllo sui dati personali o, in tempi più recenti, il c.d. diritto all'oblio. Le questioni sollevate dall'intelligenza artificiale, però, paiono diverse dalle precedenti, perché non riguardano la libertà di manifestare all'esterno, o celare ad occhi indiscreti, la propria personalità, ma quella, antecedente, di formare quest'ultima senza indebite pressioni. Per quanto, infatti, lo sviluppo dell'identità della persona non avvenga mai in modo del tutto libero – le influenze di famiglia, istituzioni e ambiente sono innumerevoli – alcuni utilizzi dell'intelligenza artificiale, *in primis* il suo impiego per l'analisi di dati personali diffusi dagli utenti nel c.d. Internet 2.0, o raccolti attraverso tecnologie IoT, rappresentano una forma di condizionamento dell'individuo inedita ed estremamente penetrante. L'elaborazione automatizzata di una mole di informazioni impensabile fino a pochi decenni fa permette di predire con livelli di accuratezza che rasentano, talvolta, la perfezione gusti, preferenze ed opinioni dell'individuo, fino a dischiudere la possibilità di influenzare questi ultimi, e i comportamenti che ne derivano. L'analisi delle abitudini di acquisto di un utente di un sito di *e-commerce*, al fine di personalizzare i prodotti che gli vengono proposti, è un ovvio esempio di questa nuova, digitale applicazione della *nudge theory*. L'esposizione prolungata, sui *social network* e altri servizi di internet interattivo, di un alto numero di utenti a contenuti elaborati in base alle loro preferenze, al fine di indirizzarne il comportamento, ha già portato a esiti preoccupanti: si pensi alla forte polarizzazione politica registrata in corrispondenza di diversi recenti appuntamenti elettorali, sfociata, nel caso delle ultime elezioni presidenziali negli Stati Uniti, in un atto apertamente eversivo come il tentato assalto al Campidoglio del 6 gennaio 2021. Più in generale, l'uso massiccio di tecniche di profilazione *online* – peraltro destinato, con la diffusione dell'IoT, a invadere sempre di più anche la vita *offline* – pare in grado di esporre i cittadini a un perpetuo *bias* di conferma, in grado di inibire le evoluzioni e i mutamenti di prospettiva che rappresentano uno degli elementi essenziali del «libero sviluppo della personalità» cui fa riferimento anche la nostra Carta fondamentale. Una volta passate in rassegna queste criticità, il lavoro esamina, in una prospettiva *de iure condendo*, i possibili strumenti giuridici innovativi con cui preservare l'effettività dei diritti afferenti alla sfera dell'identità personale.

In secondo luogo, la seconda parte del lavoro approfondisce l'impatto dell'intelligenza artificiale sulla libertà di manifestazione del pensiero. La comparsa delle prime forme di internet interattivo,

basate sulla diffusione di contenuti generati o condivisi dagli utenti, è stata salutata, due decenni fa, come l'avvio di un'era di inedite possibilità di espressione per l'uomo comune, in grado di raggiungere una platea di utenti inimmaginabile in epoca pre-digitale. Lo sviluppo successivo dei social network, però, ha mostrato anche un volto diverso: gli utenti con maggiori disponibilità economiche si sono dimostrati in grado di influenzare pesantemente il discorso pubblico al loro interno, sfruttando strumentalmente i meccanismi della viralità; il mercato delle reti sociali ha subito un rapido e vertiginoso accentramento, riducendosi all'oligopolio di una manciata di operatori; la diffusione ripetuta di notizie false e contenuti disturbanti – o, in molti paesi, apertamente illegali – ha incrinato i tradizionali meccanismi di credibilità che caratterizzavano il mercato dell'informazione. La necessità di introdurre forme di regolazione e moderazione dei contenuti diffusi su tali servizi è parsa sempre più evidente, ma i criteri adottati dalle società che li gestiscono sono spesso sembrati oscuri e contraddittori. La situazione, inoltre, ha messo in luce un nuovo problema: la concentrazione di poteri sostanzialmente censori nelle mani di pochi e potentissimi operatori privati, che appare molto complicato sottoporre a un effettivo controllo dei poteri pubblici. Il lavoro approfondisce, prima di tutto, il ruolo dell'intelligenza artificiale nei sistemi di moderazione messi a punto dalle piattaforme: il volume dei contenuti coinvolti, infatti, rende irrinunciabile l'utilizzo di filtri automatizzati. Si tratta di tecnologie di varia natura, riconducibili ad approcci all'intelligenza artificiale anche estremamente diversificati, accomunate dal garantire, allo stato dell'arte, risultati accurati, ma non perfetti. Il coinvolgimento dell'essere umano nella moderazione non pare mettere al riparo dal rischio di errori e violazioni ingiustificate della libertà d'espressione di alcuni soggetti. Anzi, le condizioni dei lavoratori impiegati nell'attività di moderazione dai giganti di internet sono spesso state descritte come critiche: individui costretti a prendere decisioni complesse in pochi secondi, estremamente sottopagati, segnati dagli effetti psicologici generati dalla prolungata esposizione a contenuti spesso crudi, violenti, moralmente sconvolgenti. In parallelo, sono esaminati alcuni dei più rilevanti casi di *deplatforming* avvenuti negli ultimi anni (si pensi, ad esempio, alla contrapposizione tra l'ex presidente statunitense Trump e i principali *social network*) e le soluzioni elaborate in materia dalle corti di alcuni Paesi, chiamate a decidere su di tali vicende. L'analisi mirerà ad evidenziare come l'attuale inquadramento giuridico delle piattaforme come comuni società di diritto privato risulti insoddisfacente, alla luce del ruolo che esse hanno assunto nel *marketplace of ideas*. A partire da questa constatazione, la ricerca esamina l'attuale assetto della regolazione degli intermediari di internet, com'è noto basato sull'esenzione da una responsabilità generalizzata per i materiali condivisi attraverso di essi. Il lavoro indaga le possibili strategie per adeguare il contesto normativo esistente al mutamento di scenario finora descritto, al fine di sviluppare un quadro giuridico in

grado, da un lato, di garantire l'effettività della libertà di manifestazione del pensiero, e, dall'altro, di proteggere l'ordine pubblico, e la stessa tenuta del sistema democratico, dai pericoli che la manipolazione del discorso pubblico sui *social network* ha dimostrato di poter generare.

Da ultimo, la seconda parte dell'opera prende in considerazione l'effetto delle tecnologie intelligenti sulla realizzazione del principio di eguaglianza. Al centro della trattazione è posto il concetto di *bias* algoritmico, oggetto di crescente attenzione da parte della letteratura scientifica. Il termine identifica ogni forma di discriminazione generata dall'utilizzo, in processi decisionali, di tecnologie intelligenti, generalmente basate sull'analisi dei dati con strumenti di apprendimento automatico. Tali effetti discriminatori possono essere causati da molteplici fattori, dal semplice errore umano nella progettazione, alla presenza, nei dati di addestramento dell'algoritmo, di differenziazioni ingiustificate a svantaggio di talune categorie, specchio di diseguaglianze esistenti nella società. Fatalmente, un sistema intelligente "allenato" con informazioni relative a una società diseguale finirà per replicare, con la precisione che ne caratterizza i risultati, le discriminazioni che la caratterizzano. Le quali, peraltro, potrebbero risultare particolarmente difficili da individuare, poiché coperte dalla percezione di oggettività che sovente, nel senso comune, circonda la tecnologia.

Il lavoro analizzerà diversi casi di *bias* algoritmico che hanno acquisito particolare notorietà, e portato i giudici di alcuni paesi ad esaminare vicende di discriminazioni generate dalla tecnologia che hanno coinvolto diritti fondamentali basilari, come la libertà personale di indagati e imputati in procedimenti penali. Successivamente, la trattazione si concentrerà sulle possibili strategie normative per limitare il fenomeno della discriminazione algoritmica, le cui conseguenze, in futuro, sembrano potenzialmente nefaste, alla luce del crescente coinvolgimento di sistemi intelligenti nei processi decisionali più vari. Nello specifico, l'analisi prenderà in considerazione l'ipotesi di introdurre nuove norme a tutela dell'uguaglianza, allargando il *corpus* del c.d. diritto antidiscriminatorio, e l'opzione - che pare, in prospettiva, decisamente più efficace - di intervenire con norme tecniche in grado di garantire che gli algoritmi siano sviluppati e allenati con *dataset* di qualità, in grado di limitare al massimo i potenziali effetti discriminatori.

La terza e ultima parte dell'opera, invece, è dedicata alla possibilità di arricchire, a causa della rivoluzione connessa all'intelligenza artificiale, il catalogo di diritti patrimonio della nostra tradizione costituzionale con nuove posizioni giuridiche degne di tutela al massimo livello. Il tema della configurabilità di c.d. *nuovi diritti* è estremamente noto e risalente: nell'ordinamento italiano è collegato al dibattito sulla corretta interpretazione dell'art. 2 della Carta, come clausola di apertura al riconoscimento di diritti in essa non enumerati, o semplice norma di rinvio alle istanze individuali di cui la Costituzione esplicita la protezione. A prevalere è stata la tesi della "fattispecie aperta",

sposata anche dalla Corte costituzionale; il tema della riconoscibilità di nuovi diritti, peraltro, trascende il diritto costituzionale, ed è patrimonio anche del diritto internazionale (si pensi alla nota ricostruzione dei diritti umani come un fenomeno caratterizzato da "generazioni" successive). La terza parte del lavoro sarà dedicata all'approfondimento di una ben precisa posizione teorica: le trasformazioni causate dalle tecnologie intelligenti sono d'intensità tale da non rappresentare solo una sfida inedita per la tutela di una vasta gamma di diritti consolidati, ma richiedono lo sviluppo di forme di protezione dell'individuo del tutto nuove, d'importanza tale da rientrare nella categoria dei diritti fondamentali. Nello specifico, saranno analizzati tre nuovi diritti la cui tutela sembra essere necessaria nell'era dell'intelligenza artificiale: il diritto di conoscere, in ogni circostanza, la natura artificiale di un sistema; il diritto a una spiegazione degli output delle tecnologie intelligenti; il diritto a un controllo umano su queste ultime. La scelta è caduta su queste tre situazioni giuridiche, tra le varie, innovative forme di tutela ipotizzate in dottrina (non sempre, in ogni caso, qualificate come diritti fondamentali) perché esse paiono poter costituire, più di ogni altra, un presidio efficace di fronte al rischio che distingue l'intelligenza artificiale da ogni innovazione tecnologica che l'ha preceduta: sovvertire la concezione antropocentrica della realtà che connota l'essere umano. L'intelligenza artificiale, infatti, priva l'uomo dell'esclusiva sulle abilità che, anche dopo le rivoluzioni industriale e digitale, sembravano caratterizzarne l'unicità: ragionamento analogico e astratto, creatività, intuito, valutazioni probabilistiche ed euristiche. Il pericolo di un'eventuale perdita del controllo su tali tecnologie non è rappresentato da scenari fantascientifici, in cui robot dominano il mondo, ma dal non riconoscere più la presenza, il reale meccanismo di funzionamento e i limiti di tali tecnologie, finendo per risultare incapaci di valutarne e, se ritenuto opportuno, sovvertirne i risultati. Una situazione che si rivelerebbe, in estrema sintesi, un'inedita forma di reificazione dell'essere umano, ridotto a oggetto destinato a subire le conseguenze di un'innovazione tecnologica che non governa più: un risultato ovviamente incompatibile con la difesa della centralità dell'individuo alla base dell'intera teorica dei diritti fondamentali, da scongiurare, dunque, con l'introduzione di nuove tutele di rango primario.

I tre diritti sono analizzati in profondità, esaminando gli sviluppi tecnologici che ne rendono necessaria la teorizzazione, le possibili aree di applicazione e le prime ipotesi, già codificate o in fase di discussione, di un loro riconoscimento legislativo. Il diritto di conoscere la natura artificiale di un sistema rimonta alle origini dell'intelligenza artificiale – Alan Turing, nel suo celebre test, concepisce l'indistinguibilità dall'essere umano come l'elemento decisivo perché una macchina possa dirsi *pensante* – e ha un'ovvia funzione antropocentrica, mirando a garantire che la riconoscibilità dell'intelligenza artificiale sia, in ogni caso, immediata (si pensi all'utilizzo di *bot* per determinate comunicazioni, o alla diffusione di contenuti *deepfake*).

Il diritto a una spiegazione si riconnette a una delle caratteristiche di determinate tecnologie intelligenti che, negli ultimi anni, ha attirato maggiormente l'attenzione della letteratura specialistica, etica e giuridica: alcune applicazioni dell'apprendimento automatico, in particolare quelle basate sulle reti neurali profonde, risultano difficilmente comprensibili dall'esterno. Ciò a causa dell'estrema complessità dei loro stati interni e del loro funzionamento basato sull'analisi di grandi moli di dati, alla ricerca di correlazioni che, per quanto accurate, rimangono meramente statistiche e in seno alle quali non risulta sempre identificabile un nesso di causalità. Ne deriva che, come detto, è particolarmente complesso - per un osservatore esterno, ma anche per lo stesso sviluppatore dell'algoritmo - ricostruire nei termini di un discorso logico-deduttivo il percorso che conduce dall'*input* all'*output* del sistema. Come si vedrà, inquadrare la spiegazione come diritto fondamentale significa mettere l'essere umano al riparo dal rischio di interfacciarsi con tecnologie che non può comprendere e, dunque, per definizione ingovernabili. In ogni caso, uno degli obiettivi essenziali del lavoro sarà evidenziare come il problema non sia risolvibile in termini netti, adottando un approccio binario in base al quale una tecnologia è spiegabile o non lo è. Gli approcci all'intelligenza artificiale meno interpretabili, infatti, garantiscono, in alcuni casi, prestazioni non ottenibili con tecnologie d'altra natura: sarà l'assetto di interessi di volta in volta in gioco, allora, a definire quali siano, in concreto, il livello di opacità e la tipologia di spiegazione accettabili, in funzione del bilanciamento con gli altri beni giuridici coinvolti. Solo se interpretato in questo modo, il diritto alla spiegazione potrà rappresentare un incentivo allo sviluppo del settore della c.d. *explainable artificial intelligence* e non un freno all'utilizzo di tecnologie che, allo stato dell'arte, appaiono in alcuni casi irrinunciabili.

Il diritto al controllo umano sul sistema consiste nella capacità di influire sul funzionamento di quest'ultimo, ad esempio condizionandone i risultati, ignorandoli o sovvertendoli completamente. La ricerca analizzerà nel dettaglio i distinti livelli di controllo umano teorizzati dalla letteratura scientifica, al fine di evidenziarne la natura di presidio antropocentrico. Solo se è garantita, in ogni caso, una soglia minima di controllo e sorveglianza umani sulla tecnologia, infatti, è allontanato il rischio di una sua deriva che possa rivelarsi dannosa, o comunque non compatibile con gli obiettivi per cui è stata progettata. A questo proposito, tale soglia minima – definibile il nucleo essenziale del diritto – dovrebbe consistere, almeno, nella presenza di uno *stop button*, ovvero la possibilità di interrompere, in ogni caso, il funzionamento della macchina, e nella possibilità di non considerarne gli *output*.

Svolto l'inquadramento dei tre nuovi diritti appena descritto, la terza e ultima parte del lavoro tenta di dare concretezza all'analisi, calando tali posizioni giuridiche in tre ambiti di applicazione privilegiati: la pubblica amministrazione, il sistema giustizia e l'arte medica. La scelta dei tre settori

è avvenuta per l'importanza dei diritti coinvolti, l'apertura all'innovazione tecnologica (nel caso dell'arte medica) e la particolare evidenza con cui si svolge, al loro interno, il rapporto tra individuo e potere, fulcro di ogni diritto fondamentale (nel caso della pubblica amministrazione e della giustizia). Inizialmente, la ricerca passa in rassegna, per ciascuno dei tre ambiti, le principali ipotesi di applicazione dell'intelligenza artificiale che li caratterizzano, e le possibilità di sviluppo futuro più rilevanti. Successivamente, l'analisi cerca di misurare la tenuta dei tre nuovi diritti teorizzati nel corso del lavoro, ipotizzando la loro applicazione in fattispecie ipotetiche o realmente accadute, talvolta già oggetto di decisioni dei giudici di alcuni paesi, e le esigenze di bilanciamento con le altre situazioni giuridiche coinvolte nel settore di riferimento. Questa disamina prende in considerazione, ovviamente, le ipotesi già esistenti di parziale positivizzazione delle tre nuove situazioni giuridiche e quelle attualmente oggetto di discussione, in primo luogo sullo scenario europeo. Il lavoro si conclude con l'analisi di una possibilità che ha attirato l'attenzione della comunità scientifica nel corso della pandemia di Covid-19, paradigmatica dell'intreccio di questioni etiche, antropologiche e afferenti ai diritti fondamentali – vecchi e nuovi – generato dall'avvento dell'intelligenza artificiale: l'utilizzo di sistemi di apprendimento automatico a supporto delle scelte tragiche in ambito sanitario.

Volgendo al termine di questa breve introduzione, pare doverosa una considerazione metodologica, riguardante il ruolo della comparazione giuridica nella ricerca. La categoria dei diritti fondamentali si presta particolarmente all'analisi comparata, a causa delle profonde similitudini, in materia, tra distinti ordinamenti, in particolare nella tradizione giuridica occidentale, e delle rilevanti differenze nella loro concezione comunque esistenti in seno a tale tradizione giuridica, in primo luogo tra il sistema statunitense e le principali democrazie europee. Tali distinzioni sono parse di particolare rilievo in alcuni punti del lavoro, in cui sono state analizzate in profondità. Si tratta, in particolare, delle parti della ricerca dedicate allo studio delle conseguenze in materia di diritti fondamentali dell'esercizio di funzioni tradizionalmente pubblicistiche (si pensi agli aspetti censori della *content moderation* nelle reti sociali) da parte di potenti operatori privati. Le differenze che caratterizzano Europa e Stati Uniti riguardo alla possibilità di un'applicazione orizzontale dei diritti fondamentali, in grado di vincolare direttamente attori privati, sono, infatti, ben note.

Allo stesso tempo, preme evidenziare che le ipotesi di regolazione dell'intelligenza artificiale, o, comunque, le norme - di *hard law* e *soft law* – e le elaborazioni giurisprudenziali rilevanti, anche indirettamente, per la materia, non sono molto numerose, spesso riguardano ambiti ben delimitati e non sono distribuite in modo uniforme tra i diversi ordinamenti. Per questa ragione, la ricerca non può dirsi uno studio che porta avanti una comparazione, sistematica e organizzata, tra due o più ordinamenti nel corso dell'intero lavoro. È parso più conveniente, infatti, trattare in modo puntuale i

singoli ordinamenti che apparissero, in un momento specifico, di particolare rilievo per lo studio di un determinato diritto fondamentale. È possibile, comunque, identificare alcuni tratti caratterizzanti dell'utilizzo degli ordinamenti stranieri fatto nel corso dell'intero lavoro: in primo luogo, lo studio prende in considerazione solo sistemi democratici, com'è inevitabile per una ricerca sulle strategie di tutela di alcuni diritti fondamentali, un tema che impone di prendere in considerazione solo ordinamenti fondati sulla garanzia dei diritti; in secondo luogo, la tradizione costituzionale italiana e l'interpretazione dei diritti data dall'ordinamento dell'Unione europea rappresentano, pressoché in ogni punto della trattazione, il riferimento principale; infine, la comparazione con gli ordinamenti di Canada e Stati Uniti (tenendo conto sia della dimensione federale che delle peculiarità di alcuni stati federati, puntualmente descritte) è quella che, nel corso dell'opera, risulta statisticamente più frequente.

In questo quadro, si inserisce la Proposta di Regolamento sull'intelligenza artificiale presentata dalla Commissione Europea il 21 aprile 2021, quando le ricerche alla base di questo lavoro erano già entrate nel vivo. Allo stato dell'arte, l'atto legislativo rappresenta l'unica ipotesi di regolazione generale dell'intelligenza artificiale, destinata, in caso di applicazione, a trovare nel Regolamento una disciplina di principio completa, dettata per ogni tipologia di tecnologia e campo di applicazione. Come si vedrà, diverse parti dell'atto sono analizzate nel dettaglio nel corso dell'opera: le soluzioni da esso proposte sono spesso presentate come possibili soluzioni alle criticità per la tutela dei diritti fondamentali evidenziate nel lavoro, o come ipotesi di possibile positivizzazione futura dei nuovi diritti qui teorizzati. Inoltre, il Regolamento è, chiaramente, uno dei principali punti di partenza delle considerazioni di diritto comparato descritte poco sopra. Ciò nonostante, il lavoro non è interamente dedicato alla Proposta della Commissione, e ampie parti di essa, considerate di scarso interesse ai fini della ricerca, non sono state trattate. In breve, i protagonisti principali dell'opera rimangono, in ogni momento, i diritti fondamentali, ed è in base a questa chiave di lettura che si auspica che essa venga letta.

PARTE I

Intelligenza artificiale: storia e caratteristiche, dilemmi etici e tentativi di regolazione

La definizione di intelligenza artificiale

1. Una pluralità di applicazioni e definizioni

Di intelligenza artificiale (da qui in avanti, spesso indicata anche con la sigla IA), il principale oggetto di studio di questo lavoro, non è agevole dare una definizione che metta al riparoda critiche di incompletezza o, all'opposto, eccessiva vaghezza e genericità. La stessa letteratura scientifica di settore è pressoché concorde nel ritenere che non ne esista una definizione universalmente riconosciuta[1]. Infatti, nel corso dei decenni l'espressione è stata usata da scienziati e ricercatori per riferirsi a tecnologie molto diverse, con premesse di partenza e obiettivi spesso divergenti. Basti pensare al variegato elenco di applicazioni – cui potrebbero, in verità, aggiungersene altre[2] – indicate dagli informatici statunitensi Stuart Russell e Peter Norvig come le più rappresentative nel principale manuale introduttivo alla materia, il loro *Artificial Intelligence - A Modern Approach*: robotica, computer vision, machine learning, ragionamento automatico,knowledge representation ed elaborazione del linguaggio naturale[3]. Non deve sorprendere, allora, che diversi specialisti considerino il termine problematico e preferiscano identificarsi come tecnici del sottosettore di riferimento[4]. Peraltro, di recente l'ambiguità semantica è ulteriormente aumentata, in ragione dell'entusiasmo suscitato dai rapidi progressi tecnologici nel campo[5] e dell'attenzione mediatica

[1] S. RUSSELL, P. NORVIG, *Artificial intelligence: a modern approach (4ª ed.)*, Hoboken (NJ), 2021, p. 1-2; P. WANG, *On defining artificial intelligence*, in *Journal of General Artificial Intelligence*, 10, 2, 2019, p. 1-37; D. MONETT, C.W.P. LEWIS, *Getting clarity by defining Artificial Intelligence - A Survey*, in V.C. MULLER (ED.), *Philosophy and Theory of Artificial Intelligence*, Berlino, 2017, p. 212–214; N.J. NILSSON, *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, Cambridge, 2009, p. XIII-XIV; R.J. BRACHMAN, *(AA)AI — more than the sum of its parts - 2005 AAAI Presidential Address*, in *AI Magazine*, 27, 4, 2006, p. 19–34.

[2] Si veda, ad esempio, J. MCCARTHY, *What is artificial intelligence?*, Stanford, 2007, http://jmc.stanford.edu/articles/whatisai.html (22/01/2021), che individua i principali settori dell'IA in: *«logical AI, search AI, pattern recognition, representation, inference, common sense knowledge and reasoning, learning from experience, planning, epistemology, ontology, heuristics, genetic programming»* ed enumera, come applicazioni più significative *«game playing, speech recognition, understanding natural language, computer vision, expert systems, heuristic classification»*.

[3] S. RUSSELL, P. NORVIG, *Artificial intelligence cit. (4ªed.)*, p. 2.

[4] AI FOR GOOD GLOBAL SUMMIT 2017, *Report*, Ginevra, 7-9 giugno 2017, p. 13-14, https://bit.ly/2ZRbfAp.

[5] Si pensi, ad esempio, ai progressi ottenuti nell'ultimo decennio nell'ambito dei veicoli a guida autonoma, della traduzione automatica, degli assistenti digitali o dei device di supporto all'attività diagnostica e chirurgica, v. L.

verso i successi ottenuti da sistemi intelligenti contro i campioni di alcuni giochi di strategia altamente rappresentativi dell'intelligenza umana [6]. Correttamente, diversi osservatori hanno evidenziato la tendenza a etichettare come applicazioni dell'intelligenza artificiale prodotti che tali non sono, o lo sono in una minima parte delle loro componenti, al fine di sfruttare commercialmente questo picco d'interesse[7].

Fermi questi limiti, è indubbio che l'elaborazione di una definizione generale di intelligenza artificiale conservi una grande utilità e, del resto, se così non fosse non si giustificherebbero i notevoli sforzi in tal senso, anche a livello istituzionale[8]. Sul punto, va evidenziato che, negli ultimi anni– in particolare, come si vedrà, in seguito agli sviluppi del c.d. *deep learning*[9]– si è registrata una crescente attenzione da parte di *policymaker* nazionali e internazionali verso questo insieme di tecnologie. Infatti, la crescente presenza di sistemi intelligenti nella vita quotidiana delle persone comuni, spesso inconsapevoli, rende ogni giorno più urgente lo sviluppo, a vari livelli, di sistemi di regolazione, di cui diviene necessario delimitare l'oggetto. Al netto delle diversità che emergono nel dibattito tecnico-scientifico, è stato correttamente evidenziato che l'elaborazione di una definizione convenzionale è necessaria per ogni tentativo di regolazione, anche al solo fine di individuare gli sviluppi tecnologici più desiderabili e tentare di orientare la ricerca e il mercato verso di essi[10].

2. La definizione di intelligenza artificiale nella letteratura scientifica: l'affermarsi dell'idea di agente razionale

Le varie definizioni di intelligenza artificiale formulate dagli studiosi che hanno contribuito alla nascita e allo sviluppo della disciplina possono raggrupparsi in quattro categorie, ciascuna rappresentante un diverso paradigma filosofico e concettuale di partenza[11]. Questa suddivisione di

---

DORMEHL, *Revisiting the rise of A.I.: How far hasartificial intelligence come since2010?,* in *Digitaltrends*, 2019, https://bit.ly/2QNy1Ew (21 gennaio 2021).

[6] Possono menzionarsi: la vittoria a Go del programma di Google DeepMindAlphaGo contro il campione sudcoreano Lee Sedol, nel 2016; la vittoria, nel 2011, del sistema esperto IBM Watson contro Ken Jennings e Brad Rutter, due dei migliori concorrenti del *game show* statunitense Jepoardy; la sconfitta, risalente al 1997, del maestro di scacchi russo Garry Kasparov contro IBM Deep Blue; la vittoria del programma Pluribus in un torneo di poker contro alcuni dei migliori giocatori professionisti del mondo.Cfr. B. MARR, *Man vs. machine: the 6 greatest AI challenge to showcase thepower of artificial intelligence*, in *Forbes (online)*, 2019, https://bit.ly/3hH5TPm (21 gennaio 2021).

[7] B. KARDON, *Is every company really an AI company?* in *AdAge*, 2019, http://bit.ly/3gNevm6 (21gennaio2021); G. ROBERTSON, *Is artificial intelligence (AI) just a buzzword?* in *Speechmatics*, 2019, https://bit.ly/2QETXBJ (21gennaio 2021).

[8] Il riferimento immediato è al lavoro dell'HIGH LEVEL EXPERT GROUP ON AI della Commissione UE, *A definition of Artificial Intelligence: maincapabilities and scientific discipline*, 8 aprile 2019.

[9] P. WANG, *On defining artificial intelligence cit.*, p. 2.

[10] Cfr. S. BHATNAGAR E AL., *Mapping intelligence: requirements and possibilities*, in V.C. MULLER (ED.), *Philosophy and Theory of Artificial Intelligence*, Berlino, 2017, p. 118: «it is difficult for policy makers to assess what AI systems will be able to do in the near future, and how the field may get there. There is no common framework to determine which kinds of AI systems are even desirable»; v. ancoraP. WANG, *On defining artificial intelligence cit.*, p. 2.

[11] Per questa classificazione cfr. S. RUSSELL, P. NORVIG, *Artificial intelligence cit.(4ª ed.)*, p. 1 ss.

massima deriva, in realtà, dalle differenze tra i vari filoni di ricerca su due temi fondamentali: il concetto di intelligenza e il rapporto tra ragionamento e comportamento. Riguardo al primo tema, lo scenario si è diviso tra chi ritiene che l'obiettivo dell'intelligenza artificiale sia replicare le abilità cognitive umane e chi, invece, si rifà a un canone astratto di *razionalità*, nella convinzione che quella umana non sia l'unica forma d'intelligenza possibile e l'IA non debba necessariamente replicarla. In merito al secondo, alcuni hanno individuato l'obiettivo dell'IA nel replicare il funzionamento del ragionamento (percepire, dedurre, apprendere, inferire, ecc.) mentre altri hanno posto l'enfasi sul comportamento intelligente, dandosi il fine di sviluppare agenti artificiali in grado di svolgere funzioni in autonomia, anche rispondendo agli input dell'ambiente in cui operano.

La prima concezione di intelligenza artificiale ad acquisire importanza nel dibattito scientifico assumeva a modello l'essere umano, puntando a imitarne sia l'intelligenza che il comportamento[12]. Il nucleo centrale dell'idea può attribuirsi al matematico britannico Alan Turing, che, nel celebre articolo *Computing Machinery and Intelligence*, pubblicato sulla rivista *Mind* nell'ottobre 1950[13], ha proposto un test al superamento del quale si sarebbe potuto dire che una macchina "pensasse"[14]. Il procedimento, chiamato dall'autore *The Imitation Game* e poi passato alla storia, appunto, come *test di Turing*, prevede che un essere umano si interfacci contemporaneamente con un suo simile e un computer, comunicando da luoghi diversi con messaggi di testo. La macchina supera il test, e può quindi considerarsi "pensante", quando l'operatore umano non è in grado di distinguere quale dei due interlocutori sia artificiale[15]. Scopo dell'intelligenza artificiale, allora, è *agire come un*

[12]Su questa linea, sono state proposte definizioni di intelligenza artificiale come «the art of creating machines that perform functions that require intelligence when performed by people» o «the study of how to make computers do things at which, at the moment, people are better», rispettivamente da R. KURZWEIL, *The age of intelligence machines*, Cambridge (USA), 1990 e E. RICH, K. KNIGHT, *Artificial intelligence (2 ed.)*, New York, 1991, p. 3.

[13] A. TURING, *Computing machinery and intelligence*, in *Mind*, 59, 236, 1950, p. 433-460.

[14] È doverosa una precisazione sull'utilizzo del termine "pensiero". L'articolo di Turing prende esplicitamente le mosse dalla domanda «Can machines think?», ma l'Autore non scioglie la questione dell'effettiva natura dell'attività della macchina che teorizza, considerando irrilevante definire se essa sia pensiero o solo un'imitazione di quest'ultimo: «I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so wellthat an average interrogator willnothave more than 70 per cent chance of making the right identification after five minutes of questioning. The original question, "Can machines think?" I believe to be too meaningless to deserve discussion».

[15] Il test di Turing è stato criticato da svariate prospettive, generalmente con l'argomento che quella esibita dal computer che teorizza non sarebbe, comunque, intelligenza. Lo stesso Autore, nel citato articolo *Computing machinery and intelligence*, elenca alcune delle contestazioni più comuni a quanto afferma, proponendo una confutazione per ciascuna di esse. È particolarmente nota la critica c.d. della *stanza cinese*, formulata dal filosofo statunitense John Searle in J.R. SEARLE, *Minds, brains and programs*, in *The behavioral and brain sciences*, 3, 1980, p. 417-424, v. anche J.R. SEARLE, *Is the brain's mind a computer program?,* in *Scientific American*, Jan. 1990, 262, 1, p. 26-31. L'argomento si basa sulla distinzione tra sintassi e semantica. Searle ritiene che la macchina ipotizzata da Turing non possa, in ogni caso, paragonarsi a una mente, poiché dalle regole che applica per comunicare in forma scritta con l'interrogante non deriverebbe in alcun modo la comprensione del significato di tali azioni e di quanto scrive. L'Autore equipara il computer di Turing a un umano non sinofono cui sia fornito un manuale con ogni regola necessaria per rispondere in forma scritta, attraverso la sola manipolazione di simboli grafici, a messaggi che gli vengono inviati con ideogrammi cinesi di cui non comprende il significato. Teorizzando la possibilità di codificare un numero sufficientemente ampio di regole da far fronte alla varietà delle forme di espressione linguistica, l'uomo potrebbe efficacemente comunicare in cinese, senza per questo conoscere la lingua e, appunto, comprendere il senso di ciò che fa. L'argomento della stanza

essere umano, fino al punto di rendersi indistinguibile da quest'ultimo. Il test originale esclude la dimensione della corporeità, stabilendo che la macchina debba riuscire a comportarsi come un essere umano soltanto dal punto di vista comunicativo. A questo proposito, una versione più estesa del test, detta *Total Turing Test*, è stata successivamente elaborata dallo scienziato cognitivo ungherese Stevan Robert Harnad[16]. Il Total Turing Test contempla la possibilità, per l'umano, di verificare anche le abilità percettive e di movimento della macchina (che deve includere tecnologie di computer vision e robotica) le quali, per superare il test, devono essere equivalenti a quelle umane[17]. Risulta ancora più evidente, allora, come la finalità dell'intelligenza artificiale, nella concezione che ha trovato origine nel test di Turing, sia lo sviluppo di agenti autonomi in grado di replicare quanto più fedelmente possibile intelligenza e comportamento umani.Questa impostazione, per quanto affascinante, non è stata particolarmente seguita dalla ricerca scientifica, che non ha concentrato i propri sforzi nella progettazione di computer in grado di superare il test di Turing. Infatti, è parso più utile, pur con accezioni differenti, tentare di sviluppare macchine che replichino i risultati dell'attività intelligente umana, senza porsi l'obiettivo forzato di riprodurne anche i meccanismi di funzionamento e il comportamento[18].

Una seconda concezione dell'intelligenza artificiale ha assunto come fine *pensare come un essere umano*, concentrandosi sull'imitazione dell'intelligenza e trascurando il comportamento. Ispirandosi a questo punto di vista, c'è stato chi ha proposto come definizioni dell'intelligenza artificiale formule icastiche come «make computer thinks»[19]o «machines with minds, in the full and literal

---

cinese ha suscitato un vivace dibattito (lo stesso Searle, nell'indicato saggio *Minds, brains and programs*, analizza le più comuni; v. anche P.M. CHURCHLAND, P.S. CHURCHLAND, *Could a machine think?,* in *Scientific American*, 262, 1, 1990, p. 32-39). In generale, riguardo all'intero dibattito attorno alla natura "pensante" di un sistema indistinguibile da un essere umano, dev'essere evidenziato che la stessa assunzione che gli altri esseri umani abbiano una coscienza è per ciascuno di noi indimostrabile e basata, appunto, sulla similitudine tra i loro e i nostri comportamenti. Il che riporta, in fondo, alle considerazioni di Turing sulla scarsa rilevanza della domanda "Can machines think?" riportate alla nota precedente. Per un'ampia raccolta di saggi a commento del test di Turing si rimanda a R. EPSTEIN, G. ROBERT, G. BEBER (a cura di), *Parsing the Turing Test*, Dordrecht, 2008.

[16]Tra gli altri, S. R. HARNAD, *Minds, machines and Searle*, in *Journal of theoretical and experimental artificial intelligence*, 1, 1989, p. 5-25; *Other bodies, other minds: a machine incarnation of an old philosophical problem*, in *Minds and Machines*, 1, 1991, p. 43-54; *The Turing test is not a trick: Turing indistinguishability is a scientific criterion*, in *ACM SIGART Bulletin*, 3, 4, 1992, p. 9-10.

[17]Harnad, in ogni caso, afferma che il problema della natura delle abilità dimostrate da una macchina non ha soluzione, e che non vi è modo per definire se la macchina che teorizza sia dotata di coscienza e intelligenza o solo in grado di imitare queste ultime, estendendo il ragionamento anche alla ricostruzione degli stati mentali altrui: «If you are stillnotentirelycomfortable with equatinghaving a mind with having the capacity to produce Turing-indistinguishablebodily performance, then welcome to the mind-body problem! […] No scientific answer can be expected to the question of how or why we differ from mindless bodies that simply behave exactly as if they have minds», S.R. HARNAD, *Other bodies, other minds: a machine incarnation of an old philosophical problem cit.*, p. 51-52.

[18] Cfr. S. RUSSELL, P. NORVIG, *Artificial intelligence cit. (4ª ed.)*, p. 2; Si tratta, in ogni caso, di una disaffezione solo relativa, e il test di Turing resta rivestito di notevole valore simbolico per la ricerca sull'IA, tanto che, dal 1990, esiste una competizione annuale che premia il programma che più si avvicina al suo superamento, il *Loebner Prize*, v. J. WAKEFIELD, *The hobbyists competing to make AI human*, in BBC news (online), 13 settembre 2019, https://bbc.in/3tAwgNj (6 febbraio 2020).

[19] J. HAUGELAND, *Artificial intelligence: the very idea*, Cambridge (US)-Londra, 1985, p. 2.

sense»[20]. È un approccio fortemente debitore alle scienze cognitive, un campo di studi interdisciplinare in seno al quale sono state sviluppate varie teorie sperimentali sul funzionamento della mente umana, poi utilizzate come modelli per implementare programmi per elaboratore[21]. Tra i primi e più noti rappresentanti di questa corrente possono menzionarsi gli statunitensi Allen Newell ed Herbert Simon, che, nel 1957, svilupparono un programma noto come *General Problem Solver*[22]. Come si dirà, si trattava di un sistema in grado di risolvere, sulla base di un insieme di regole di decisione formalizzate nella programmazione, quesiti di carattere logico e geometrico. Per stessa ammissione dei due scienziati, l'obiettivo dell'iniziativa non era solamente sviluppare il programma, ma anche - e soprattutto – comparare le sue strategie di soluzione con quelle della mente umana, al fine di migliorare la nostra conoscenza dei meccanismi di quest'ultima[23].

Anche questo approccio, tuttavia, si è scontrato con l'obiezione che la semplice mimica delle abilità umane ha un valore aggiunto limitato, e che la ricerca nel campo dell'IA dovrebbe piuttosto tendere a replicarne e aumentarne, dove possibile, i risultati. È emerso, così, il menzionato concetto astratto di razionalità, intesa come la performance di un agente modello, non necessariamente equivalente all'intelligenza umana[24]. In quest'ottica, quella umana è solo una delle molte versioni dell'intelligenza possibili e, magari, non la più efficiente in ogni contesto. L'obiettivo dell'intelligenza artificiale, dunque, dev'essere lo sviluppo di sistemi che aderiscano il più possibile a tale canone ideale di razionalità. Anche tra coloro che hanno aderito a questa prospettiva si è registrata una divisione tra chi ha posto l'enfasi sul ragionamento e chi sul comportamento.

La prima di tali due posizioni, il cui obiettivo può riassumersi nello sviluppo di sistemi in grado di *pensare razionalmente*[25], ha le sue radici nel campo di studi della logica. La storia del pensiero, infatti, si è occupata fin dalla filosofia antica della formalizzazione dei meccanismi del ragionamento in regole generali e astratte, avulse dal contenuto cui sono di volta in volta applicate - la razionalità, appunto - sviluppando sistemi di notazione con cui è possibile descrivere problemi e

[20] *Ibidem*.

[21] Si vedano, tra I molti, S. GURUMOORTHY, B.N. RAO, X.Z. GAO, *Cognitive sciences and artificial intelligence*, Singapore, 2018; J. FRIEDENBERG, G. SILVERMAN, *Cognitive science: an introduction to the science of the mind (3 ed.)*, Los Angeles, 2016; M.H. BICKHARD, L. TERVEEN, *Foundational issues in artificial intelligence and cognitive science: impasse and solution*, Amsterdam, 1995.

[22] A. NEWELL, J. C. SHAW, H. A. SIMON, *Report on a general problem-solving program*, 1959, https://bit.ly/2GmYirj (21gennaio 2020).

[23] «This paper reports on a computer program, called GPS-I for General Problem-Solving Program I. Construction and investigation of this program is part of a research effort by the authors to understand the information processes that underlie human intellectual, adaptive and creative abilities. […] The paper will present enough theoretical discussion of problem-solving activity so that the program can be seen as an attempt to advance our basic knowledge of intellectual activity», A. NEWELL, J. C. SHAW, H. A. SIMON, *Report on a general problem-solving program cit.*, p. 2.

[24] V. ancora S. RUSSELL, P. NORVIG, *Artificial intelligence cit.(4ª ed.)*, p. 1-5 e 39 ss.

[25] Cfr. ad esempio E. CHARNIAK, D. MCDERMOTT, *Introduction to artificial intelligence*, Boston, 1985, chedefinisconol'IA: «The study of mental faculties through the use of computational models».

relazioni tra concetti[26]. In linea di principio, è possibile sviluppare programmi per elaboratore in grado di risolvere qualunque problema dotato di soluzione e formalizzabile logicamente, in quanto tale codificabile in linguaggio-macchina[27]. L'approccio logicista all'intelligenza artificiale mira a raggiungere il menzionato standard ideale di pensiero razionale attraverso la programmazione di sistemi di questo tipo, in grado di dominare un contesto di applicazione grazie alla formalizzazione delle regole necessarie alla sua conoscenza e alla risoluzione dei problemi che vi si presentano. È un'impostazione che si è scontrata, in passato, con due limiti: la necessità di enorme capacità computazionale, al fine di implementare nelle macchine un numero molto alto di regole di giudizio, e la difficoltà diformalizzare campi del sapere nei quali sia inaccessibile una conoscenza totale del contesto di riferimento e siano necessarie valutazioni di tipo statistico e probabilistico[28].Si tratta di problemi in parte superati nel corso dell'ultimo decennio, con l'aumento esponenziale del potere computazionale dei moderni computer e l'avvento delle reti neurali (che non possono, peraltro, ricondursi strettamente alla concezione logicista)[29]. In generale, però, l'idea che l'IA debba imitare il pensiero razionale nelimita in partenza il campo d'azione alla risoluzione di problemi di natura squisitamente concettuale. L'impostazione oggi prevalente ritiene, invece, che l'IA debba puntare allo sviluppo di sistemi in grado non solo di ragionamento, ma anche di comportamento razionale, riuscendo ad agire in modo adattativo ed efficiente nel contesto in cui operano. È comune l'utilizzo dell'espressione *agente razionale*[30].

In senso lato, ogni programma per elaboratore porta a termine delle attività ed è, dunque, un agente. Nell'accezione qui considerata, è decisiva l'interazione con l'ambiente circostante: un sistema intelligente dev'essere in grado di agire autonomamente, adattarsi ai cambiamenti esterni, modificare il proprio comportamento al fine di realizzare gli obiettiviassegnati. Risultano valorizzate branche dell'intelligenza artificiale che sono ridotte ai margini qualora ci si focalizzi unicamente sul ragionamento, come robotica e computer vision. Secondo questa impostazione, il ragionamento razionale è un presupposto necessario, ma non sufficiente di un sistema intelligente: esistono situazioni e contesti che impongono forme di *agire razionale* che non si basano solo sulla corretta applicazione di inferenze logiche, a causadella scarsità dei dati di partenza o di mutamenti

[26] Possono indicarsi, tra i molti, I. M. COPI, C. COHEN, V. RODYCH, *Introduction to logic (15ª ed.)*, Londra, 2019; C. MANGIONE, S. BOZZI, *Storia della logica*, Catania, 1985; E. J. LEMMON, *Beginning logic*, Londra-Edimburgo, 1965.
[27] N. J. NILSSON, *Logic and artificial intelligence*, in *Artificial Intelligence*, 47, 1991, p. 31-56; J. MCCARTHY, *Concept of logical AI*, in J. MINKER(a cura di), *Logic-basedartificial intelligence*, Norwell, 2000, p. 37-56.
[28] N. J. NILSSON, *Artificial intelligence: a new synthesis*, Burlington, 1998, p. 301 ss.; S. RUSSELL, P. NORVIG, *Artificial intelligence: a modern approach (3ª ed.)*, Upper Saddle River, 2010, p. 4.
[29] M.ROSER, H. RITCHIE, *Technological progress*, in *OurWorldinData.org*, 2020, https://ourworldindata.org/technological-progress (1 febbraio 2021); T. HWANG, *Computational power and the social impact of artificial intelligence*, 2018, http://dx.doi.org/10.2139/ssrn.3147971 (1febbraio 2021); S. RUSSELL, P. NORVIG, *Artificial intelligence cit. (4ª ed.)*, p. 24 ss.
[30] S. RUSSEL, P. NORVIG, *Artificial intelligence cit. (4ª ed.)*, p. 3-4 e 36 ss.;

ambientali repentini[31]. L'agente razionale è definibile come l'agente che determina il proprio comportamento in modo da raggiungere il miglior risultato possibile in una data attività, anche definendolo con modalità euristiche, qualora operi in contesti incerti[32].

Come detto, la concezione dell'intelligenza artificiale basata sul comportamento razionale è oggi prevalente nella letteratura scientifica di settore[33], per due motivi principali. In primo luogo, perché l'enfasi posta sullo sviluppo di sistemi che corrispondano alla definizione di agente razionale permette di ricondurre al concetto di intelligenza artificiale la più larga varietà di applicazioni, e tiene conto della circostanza che l'inferenza logica può non essere l'unica via per raggiungere un comportamento in senso lato razionale. In secondo luogo, perché accoglie la definizione di intelligenza più ampia possibile, ancorandola a un modello astratto di razionalità e non all'imitazione delle abilità umane[34]. La fortuna di questa impostazione, peraltro, è stata tale da trascendere i confini della produzione scientifica e venir accolta in diversi documenti che hanno tentato, a vari livelli, di definire e fornire prime prospettive di regolazione dell'intelligenza artificiale.

3. Le definizioni istituzionali di intelligenza artificiale

Negli ultimi anni, diverse istituzioni dell'Unione Europea sono state coinvolte in uno sforzo congiunto per la formulazione di una definizione generale di intelligenza artificiale. L'operazione ha avuto inizio da una Comunicazione della Commissione Europea dd. 25 aprile 2018, intitolata *Artificial Intelligence for Europe*[35], in cui era enunciata una prima"definizione istituzionale" di intelligenza artificiale, destinata a fornire la base per le elaborazioni successive:

«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)»[36].

[31] Così i citati S. RUSSEL, P. NORVIG, *Artificial intelligence cit. (4ªed.)*, p.4: «There are ways of acting rationally that cannot be said to involve inference. For example, recoiling from a hot stove is a reflex action that is usually more successful than a slower action taken after careful deliberation».

[32] *Ibidem.*

[33] Cfr. ad es. M. WOOLDRIDGE, *Reasoning about rational agents*, Cambridge, 2003; i contributi raccolti in M. WOOLDRIDGE, A. RAO (A CURA DI), *Foundation of rational agency*, Dordrecht, 1999; S. RUSSELL, P. NORVIG, *Artificial intelligence cit.*; CE HIGH LEVEL EXPERT GROUP ON AI, *A definition of Artificial Intelligence cit.*

[34] V. ancora S. RUSSELL, P. NORVIG, *Artificial intelligence cit.(4ª ed.)*, p. 3-4.

[35] 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 on Artificial Intelligence for Europe, Brussels, 25 aprile 2018 COM(2018) 237 final.

[36] Communication from the Commission on Artificial Intelligence for Europe cit., p. 2.

Per la verità, appare evidente che questa prima definizione non poteva ricondursi, senza evidenti forzature, al concetto di agente razionale. La formula proposta dalla Commissione, infatti, utilizzava l'espressione *intelligent behaviour*, interpretabile come un riferimento al modello dell'intelligenza umana. Inoltre, il termine "razionale", o suoi equivalenti della stessa area semantica, non comparivano mai.

Queste caratteristiche della definizione della Commissione sono state messe in discussione ed abbandonate nelle fasi successive dell'elaborazione, cominciate con la formazione dell'High Level Expert Group on Artificial Intelligence, un organo di 52 esperti nominato dalla Commissione Europea nel giugno 2018, al termine di una selezione pubblica basata su candidature spontanee[37]. Tra i vari documenti prodottidal gruppo, ve n'è uno intitolato *A definition of AI: Maincapabilities and disciplines* [38] , elaborato parallelamente a uno studio sull'impatto etico dell'intelligenza artificiale che ha portato alla stesura delle più note *Ethics guidelines for trustworthy AI*[39]. La prima bozza, resa pubblica il 18 dicembre 2018, conteneva una definizione di intelligenza artificiale fortemente debitrice della concezione di IA ispirata all'idea di agente razionale[40] e in cui era scomparso l'ambiguo riferimento all'*intelligent behaviour*:

«Artificial intelligence (AI) refers to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to take (according to pre-defined parameters) to achieve the given goal. AI systems can also be designed to learn to adapt their behavior 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)»[41].

Questa formulazione è stata lievemente modificata dopo una consultazione pubblica di alcuni mesi, che ha coinvolto anche le menzionate*Ethics guidelines* e portato all'attuale versione:

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

[37] L'elenco completo dei membri è consultabile alla pagina: https://bit.ly/3up7noi (16 febbraio 2021).

[38] HIGH LEVEL EXPERT GROUP ON AI, *A definition of Artificial Intelligence: main capabilities and scientific discipline cit.*, 8 aprile 2019.

[39] HIGH LEVEL EXPERT GROUP ON AI, *Ethics Guidelines for Trustworthy Artificial Intelligence*, 8 aprile 2019, https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html (7settembre 2021).

[40] Sia la prima bozza del documento che la versione finale, inoltre, enunciano, nella prima pagina: «An AI system isthus first and foremostrational».

[41] HIGH LEVEL EXPERT GROUP ON AI, *A definition of AI: main capabilities and scientific disciplines – first draft*, 18 dicembre 2018, p. 7.

environment through data acquisition, 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)»[42].

A questa definizione deve aggiungersi quella prevista dall'art. 3 della Proposta di Regolamento in materia di IA presentata dalla Commissione Europea nell'aprile 2021[43], destinata a divenire, in caso di definitiva approvazione, una sorta di "definizione ufficiale" di intelligenza artificiale, legalmente vincolante sullo scenario europeo: *«'artificial intelligence system' (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»*[44]. La formula pare riprendere, in modo molto più sintetico, i tratti essenziali di quelle elaborate dal Gruppo di Esperti. La sua principale particolarità consiste nel non menzionare sistemi *hardware*, limitando il concetto di intelligenza artificiale al solo *software*: è chiaro, comunque, che le applicazioni dell'intelligenza artificiale in cui l'*hardware* pare prevalente (la robotica, ad esempio) sono tali, in realtà, per la natura del *software* che ne determina il funzionamento, non essendo altrimenti distinguibili da ogni altra macchina. Tale differenza rispetto alle definizioni del Gruppo di Esperti non pare, dunque, particolarmente significativa.

L'Unione Europea, in ogni caso, non è l'unica organizzazione internazionale ad aver proposto una definizione istituzionale di intelligenza artificiale. Formule definitorie, infatti, sono state elaborate da diversi altri enti, spesso in documenti – tutti, allo stato, inidonei a impegnarne in modo

[42] HIGH LEVEL EXPERT GROUP ON AI, *Ethics Guidelines for Trustworthy Artificial Intelligence cit.*, 8 aprile 2019, p. 6.

[43] COMMISSIONE EUROPEA, *Proposta di Regolamento al Parlamento Europeo e al Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (Legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione*, 21 aprile 2021, COM(2021) 206 final. Le tecniche per lo sviluppo di sistemi di intelligenzaartificiale indicate dall'Annex I dellaProposta, divise in trecategorie, sono: «(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation,inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods». La Proposta sarà estensivamente presa in esame in vari punti di questo lavoro, cfr. in primo luogo *infra*, p. 69 ss.

[44] La definizione è enunciata all'art 3 par. 1 n. 1 della Proposta di Regolamento.

vincolante gli stati membri – che affrontano alcune delle problematiche etico-sociali connesse alla diffusione dell'IA. Si tratta di definizioni caratterizzate da un orientamento marcatamente operativo, che delimitano il campo dell'IA tramite l'elencazione delle sue principali applicazioni e in cui l'influenza della teoria dell'agente razionale, pur intuibile dalle tecnologie di volta in volta menzionate, rimane sullo sfondo. Inoltre, probabilmente in ragione del tipo di documenti in cui tali definizioni sono contenute, riguardanti l'impatto etico dell'IA, le applicazioni cui si riferiscono sono, in primo luogo, quelle che appaiono più problematiche per la possibilità di utilizzi distorti ed effetti indesiderati, come i sistemi in grado di formulare predizioni e prendere decisioni.

L'UNESCO, ad esempio, nella 40ª sessione plenaria del novembre 2019, ha avviato, anche in questo caso con la nomina di un gruppo di esperti[45], un percorso biennale per la stesura di una *Recommendation on the ethics of artificial intelligence*[46], definitivamente approvata dall'Assemblea Generale nel novembre 2021. Il documentoenunciaquestadefinizione di intelligenza artificiale:

«AI systems are information-processing technologies that embody models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in real and virtual environments. AI systems are designed to operate with some aspects of autonomy by means of knowledge modelling and representation and by exploiting data and calculating correlations. AI systems may include several methods, such as but not limited to: (i) machine learning, including deep learning and reinforcement learning;(ii) machine reasoning, including planning, scheduling, knowledge representation and reasoning, search and optimization.

AI systems can be used incyber-physical systems, including the Internet-of-Things, robotic systems,social robotics and human-computer interfaces which involve control, perception, the processing of data collected by sensors, and the operation of actuators in the environment in which AI systems work»[47].

Anche l'OCSE, in una *Recommendation on artificial intelligence (AI)* adottata dal Consiglio dei Ministri degli stati membri il 22 maggio 2019[48], contenente una lista di misure e principi per assicurare lo sviluppo di un'IA affidabile e *human-centred*,ha proposto una succinta definizione di intelligenza artificiale, che pone l'accento, appunto, sui sistemi che formulano predizioni, decisioni o raccomandazioni:

[45] L'elenco completo dei membri è consultabile alla pagina: https://bit.ly/3qHA9hC (16 febbraio 2021).
[46] UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, 24 novembre 2021, 41 C/73 (Annex).
[47] UNESCO AD HOC EXPERT GROUP (AHEG), *First draft of the recommendation on the ethics of artificial intelligence cit.*, Parigi, 7 settembre 2020, p. 4.
[48] OCSE CONSIGLIO DEI MINISTRI, *Recommendation on artificial intelligence*, 22 maggio 2019, OECD/LEGAL/0449, https://bit.ly/3bdlJ21 (10 febbraio 2021).

«An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy»[49].

Infine, è doveroso evidenziare che, sul piano degli ordinamenti nazionali, l'ordinamento canadese ha codificato, nel 2019, quella che, allo stato dell'arte, è la prima definizione di IA contenuta in un testo normativo vincolante e pienamente in vigore. Si tratta, peraltro, di una definizione che si distingue nettamente da quelle formulate in ambito sovranazionale, poiché si rifà decisamente alle concezioni per le quali lo scopo dell'intelligenza artificiale èreplicare il comportamento e le abilità cognitive dell'essere umano. Dal 1 aprile di quell'anno, infatti, è in vigore una direttiva del governo canadese intitolata *Directive on automated decision-making*[50]. Il testo ha lo scopo di definire modalità e limiti di utilizzo, da parte della pubblica amministrazione, di algoritmi di decisione automatizzata e ha fornito questa definizione di intelligenza artificiale:

«Information technology that performs tasks that would ordinarily require biological brainpower to accomplish, such as making sense of spoken language, learning behaviours, or solving problems»[51].

Il documento, inoltre, propone anche una definizione di *automateddecision system*,anch'essa ispirata decisamente all'idea di imitazione delle qualità umane:

«Includes any technology that either assists or replaces the judgement of human decision-makers. These systems draw from fields like statistics, linguistics, and computer science, and use techniques such as rules-based systems, regression, predictive analytics, machine learning, deep learning, and neural nets.»[52].

4. L'intelligenza artificiale è *intelligenza*? Il problema dell'utilizzo di terminologie antropomorfe

A margine delle varie definizioni formulate nel corso del tempo, è doveroso chiedersi se la stessa espressione *intelligenza artificiale* sia la più adatta allo scopo. Infatti, non mancano le voci critiche, specialmente riguardo all'impiego del termine "intelligenza", considerato del tutto inadeguato allo stato attuale della tecnologia, e ritenuto, in ogni caso, esclusiva dell'essere umano[53]. In quest'ottica, bisognerebbe evitare di parlare di intelligenza al fine di evidenziare le radicali differenze con l'essere umano che permarrebbero anche qualora fossero sviluppate macchine dotate di intelligenza

[49] OCSE CONSIGLIO DEI MINISTRI, *Recommendation on artificial intelligence cit.*, 22 maggio 2019, p. 7.
[50] Government of Canada, *Directive on automated decision-making*, 1 aprile 2019, https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592 (7 novembre 2021).
[51] Government of Canada, *Directive on automated decision-making cit.*, Annex 1.
[52] *Ibidem.*
[53] Cfr. L. JULIA, *L'intelligence artificial n'existe pas*, Parigi, 2019; M. SIMKOFF, A. MAHDAVI, *AI doesn't actually exist yet*, in *Scientific American – Observations*, 12 novembre 2019, https://bit.ly/3nViJPs (26 novembre 2021).

generale (c.d. *strong AI*[54]). L'artificiale rimarrebbe, anche in tal caso, del tutto estraneo alla condizione umana, e i suoi comportamenti razionali non potrebbero comunque dirsi intelligenti, posto che tale termine sarebbe esclusiva delle capacità intellettive, della coscienza e della biologia dell'essere umano.

Si tratta di una tesi che ha il merito di evidenziare che la differenza ontologica tra umano e artificiale continuerebbe a esistere, inalterata, anche con lo sviluppo di macchine antropomorfe sempre più raffinate, per le quali, dal punto di vista del diritto, andrebbe probabilmente creato un autonomo status, piuttosto che tentare di adattare il quadro giuridico applicato alle persone fisiche[55].

Tuttavia, è opportuno muovervi alcune specificazioni.In primo luogo, va registrato che il termine intelligenza è intrinsecamente ambiguo. Le incertezze su cosa sia, in fondo, questa particolare qualità umana hanno attraversato la storia del pensiero[56], e non manca chi dubiti della sua stessa esistenza, almeno nel significato ad essa attribuito nella vita quotidiana[57]. In realtà, ciascun essere umano non può essere certo che i suoi simili abbiano una coscienza e degli stati mentali simili ai suoi. Lo studioso di robotica australiano Rodney Brooks ha descritto l'insondabilità degli stati mentali altrui con un efficace aforisma: «l'intelligenza è negli occhi di chi guarda»[58]. La volontà di riconoscere comportamento intelligente nei propri simili e non in un'ipotetica macchina antropomorfa in grado di replicarlo, sarebbe, dunque, frutto di una convenzione[59]. In secondo luogo, l'idea che il termine intelligenza rimandi necessariamente all'essere umano è estremamente diffusa e, come già visto, ha attraversato, tramite la dicotomia con la razionalità, la storia dell'intelligenza artificiale[60], ma rimane frutto di una forzatura semantica. Nulla vieta, infatti, di ritenere che quella dell'uomo non sia l'unica forma di intelligenza possibile, e utilizzare il termine per riferirsi a

[54] È comune distinguere tra *narrow* o *weak AI*, per riferirsi a sistemi in grado di agire in modo autonomo e adattandosi all'ambiente in campi d'azione ristretti e ben definiti, e *strong o general AI*, per riferirsi a strumenti (oggi del tutto ipotetici) in grado di operare in modo versatile in diversi campi di applicazione, analogamente all'intelligenza umana, v. J. R. SEARLE, *Minds, brains and programs cit.*, p. 417; S. BRINGSJORD, N.S. GOVINDARAJULU, *Artificial Intelligence*, in E. N. ZALTA (A CURA DI), *The Stanford Encyclopedia of Philosophy*, 2020, https://stanford.io/384A1Bl (26 novembre 2021).

[55] Si può citare, in proposito, il dibattito scaturito dalla proposta, contenuta nella Risoluzione del Parlamento Europeo del 16 febbraio 2017 recante *Raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica* (2015/2103(INL)), di considerare «l'istituzione di uno status giuridico specifico per i robot nel lungo termine». In dottrina cfr. A. SANTOSUOSSO, *The human rights of nonhuman artificial entities: an oxymoron?*, in *Jahrbuch für Wissenschaft und Ethik*, 19, 2015, pp. 203-237. L'ipotesi, in ogni caso, ha incontrato forte opposizione: si veda, ad esempio, la lettera aperta di diverse decine di esperti *Open letter to the EuropeanCommission – Artificial Intelligence and Robotics*, https://bit.ly/3q78jKA (26 febbraio 2021).

[56] Per la varietà delle possibili teorie dell'intelligenza, cfr. R. J. STERNBERG, *Intelligence*, in D. K. FREEDHEIM, I. B. WEINER (A CURA DI), *Handbook of psychology: History of psychology*, Hoboken, 2013, p. 155-176; D. WAHLSTEN, *The Theory of Biological Intelligence: History and a Critical Appraisal*, in R. J. STERNBERG, E. L. GRIGORENKO, *The general factor of intelligence*, New York, 2002.

[57] V. ad es. D. SERPICO, *Esiste davvero l'intelligenza generale? Prospettive delle scienze cognitive*, in *NeaScience*, 9, 2, 2015, p. 216-219.

[58] R.A. BROOKS, *Intelligence without reason*, in L. STEELS, R.A. BROOKS (a cura di), *The artificial life route to artificial intelligence*, Mahwah (New Jersey), 1995, p. 57.

[59] Si vedano, in generale, i noti studi di D.C. DENNET, *The intentional stance*, Cambridge, 1987.

[60] V. *supra*, p. 24 ss.

macchine che svolgono determinate operazioni in autonomia, senza per questo uscire dalla sfera dei suoi possibili significati.

Per concludere, le ambiguità connesse all'espressione "intelligenza artificiale" non paiono tali da farne abbandonare l'utilizzo, vista anche la notevole diffusione nella prassi, e in questo lavoro il termine verrà abitualmente utilizzato. Del resto, rimane di utilità relativa chiedersi se ciò che sembra intelligenza sia veramente intelligenza: è stato fatto notare che, in fondo, chiedersi se un computer pensi «non sia più utile di chiedersi se un sottomarino nuoti»[61].

[61] «The Fathers of the field had been pretty confusing: John von Neumann speculated about computers and the human brain in analogies sufficiently wild to be worthy of a medieval thinker and Alan M. Turing thought about criteria to settle the question of whether Machines Can Think, a question of which we now know that it is about as relevant as the question of whether Submarines Can Swim», E. DIJKSTRA, *The threats to computing science*, *Statement at the ACM 1984 South Central Regional Conference*, Austin, 16-18 novembre 1984, https://bit.ly/381XdjK (26febbraio2021).

Cenni storici sullo sviluppo dell'intelligenza artificiale

1. La Conferenza di Dartmouth del 1956 e le origini dell'IA

Simbolicamente, l'intelligenza artificiale ha una data di nascita: il 31 agosto 1955, data di pubblicazione della *Proposal for the Dartmouth SummerResearch Project onArtificial Intelligence*[62], un invito, rivolto a studiosi di varie discipline, a partecipare a «*a two-month, 10 men study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire*»[63]. Il documento è considerato la prima occasione in cui è stata usata l'espressione *artificial intelligence*[64]e recava le firme di John McCarthy, ideatore dell'iniziativa e al tempo matematico dell'Università di Dartmouth, ClaudeShannon, ingegnere e matematico dei Bell Telephone Laboratories, Marvin Minsky, poi cofondatore dell'*Artificial intelligence Project* al MIT,eNathaniel Rochester, informatico dell'IBM.Vi parteciparono diversi tracoloro che, in seguito, sarebbero stati i più importanti contributori allo sviluppo dell'intelligenza artificiale, come i già citati Allen Newell ed Herbert Simon, Ray Solomonoff, Arthur Samuel, Oliver Selfridge, Julian Bigelow o Donald Mac Crimmon MacKay[65]. La conferenza durò, in realtà, sei settimane, e rappresentò la prima occasione per studiosi che lavoravano, a vario titolo, a questo nuovo campo di studi di confrontarsi e riconoscersi membri della medesima comunità scientifica[66]. La proposta diMcCharty, Shannon, Minsky e Rochester indicava sette problemi, la cui soluzione interessava la nuova disciplina dell'*artificial intelligence*, con un elenco di sorprendente attualità[67]:

- come sviluppare «*automatic computers*»;
- «*how can a computer be programmed to use a language*», un settore oggi noto come *Natural Language Processing*[68];

[62] Indicano nella conferenza di Dartmouth il momento di nascita dell'IA, tra gli altri, D. CREVIER,*AI: TheTumultuos Search for Artificial Intelligence*, New York, 1993, p. 49 e S.J. RUSSELL, P. NORVIG, *Artificial Intelligence cit.(3ª ed.)*, p. 17-18. Il testo completo della proposta è disponibile sulla pagina web dedicata a uno dei partecipanti alla conferenza, lo statunitense Ray Solomonoff, v. J. MCCARTHY, M.L. MINSKY, N. ROCHESTER, C.E. SHANNON, *A proposal for the Dartmout Summer Research Project on Artificial Intelligence*, 1955, http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf (24 marzo 2021). Sull'evento di Dartmouth cfr. anche R.R. KLINE, *Cybernetics, Automata Studies, and the Dartmouth Conference on Artificial Intelligence*, in *IEEE Annals of the History of Computing*, 33, 4, 2011, p. 5-16.

[63] J. MCCARTHY, M.L. MINSKY, N. ROCHESTER, C.E. SHANNON, *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence cit.*, p. 1.

[64] P. MCCORDUCK, *Machines Who Think: a personal inquiry into the history and prospects of artificial intelligence*, Natick, 2004, p. 114-115.

[65] N. J. NILSSON, *The quest for artificial intelligence. A history of ideas and achievement*, New York, 2010, p. 53.

[66] S.J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (3ª ed.)*, p. 17-18.

[67] J. MCCARTHY, M.L. MINSKY, N. ROCHESTER, C.E. SHANNON, *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence cit.*, p. 1-3. I corsivi nell'elenco sono citazioni del testo originale.

[68] Due anni prima della conferenza di Dartmouth, nel gennaio 1954, aveva suscitato un certo clamore un esperimento organizzato congiuntamente dall'Università di Georgetown e dall'IBM, in cui un rudimentale sistema di traduzione automatica aveva tradotto dal russo all'inglese circa 60 frasi attinenti all'ambito della chimica organica, v. J. HUTCHINS,

- la creazione di «*neuron nets*» artificiali sul modello di quelle biologiche, con lo scopo di riprodurne il funzionamento;
- l'elaborazione di una «*theory of the size of a calculation*», al fine di misurare la complessità e la potenza computazionale necessaria per lo sviluppo di un calcolatore utilea risolvere un determinato problema;
- lo sviluppo di macchine capaci di «*self-improvement*»;
- la programmazione di sistemi capaci di ragionamento astratto[69];
- la creazione di macchine in grado di ragionamento creativo[70].

Si tratta di temi che erano stati, in parte, affrontati negli anni precedenti al 1956, generando il dibattito scientifico che ha poi portato alla conferenza di Dartmouth. Per ripercorrere i punti fondamentali della storia dell'intelligenza artificiale è doveroso, prima di tutto, dar conto di tali prime, pionieristiche attività.

Gli studi più risalenti, tra quelli significativi per lo sviluppo dell'intelligenza artificiale, riguardano un filone di ricerca nominato anche nella *Proposal* di Dartmouth del 1955 eche sarà, poi, per lungo tempo abbandonato a causa della mancanza dei necessari mezzi tecnici: le reti neurali. Risale addirittura al 1943, infatti, l'articolo dei due logici e neuroscienziati W.S. McCulloch e W. Pitts, *A logical calculus of the ideasimmanent in nervous activity*[71], in cui gli autori esponevano il primo modello matematico di neurone artificiale.Il neurone teorizzato da Pitts e McCullochsi sarebbe dovuto attivare, fornendo un output positivo, sulla base di stimoli forniti dagli altri neuroni cui era collegato, rimanendo in caso contrario silente, senza gradazioni intermedie. I due studiosi dimostrarono che una rete sufficientemente ampia di neuroni di quel genere avrebbe potuto eseguire operazioni e risolvere problemi basati su tutti i principali operatori logici (AND, OR, NOT, ecc.) e ipotizzarono potesse essere utilizzata per l'apprendimento automatico, come avverrà effettivamente decenni dopo[72]. Tre anni dopo D. Edmonds e lo stesso M.Minsky, all'epoca studente di matematica ad Harvard, svilupparono la prima rete neurale, formata da 40 neuroni artificiali sul modello di Pitts e McCulloch[73].

The first public demonstration of machine translation: the Georgetown-IBM system, 7th January 1954, 2006, https://bit.ly/39eWnR4 (25 marzo 2021).

[69] Il problema, nella versione originale, era presentato semplicemente con il titolo di «*Abstraction*».

[70]Nell'originale, «*Randomness and Creativity*».

[71] W.S. MCCULLOCH, W. PITTS, *A logical calculus of ideas immanent in nervous activity*, in *Bulletin of Mathematical Biophysics*, 5, 1943, p. 115-133.

[72] W.S. MCCULLOCH, W. PITTS, *A logical calculus of ideas immanent in nervous activity cit.*; S. J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (4ª ed.)*, p. 17; N.J. NILSSON, *The Quest for Artificial Intelligence: A History of Ideas and Achievements cit.*, p. 15; A. KROGH, *What are artificial neural networks?*, in *Nature Biotechnology*, 26, 2008, p. 195.

[73]D. CREVIER, *AI: The Tumultuos Search for Artificial Intelligence cit*., p. 34-35; S. J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (4ª ed.)*, p. 17.

Sempre nel 1950, Alan Turing pubblicava il menzionato articolo *Computing machinery and intelligence*[74], avviando il dibattito, destinato a durare decenni, sulla possibilità, i limiti e il significato dello sviluppo di intelligenze artificiali. Due anni dopo, nel 1952, faceva la sua comparsa un filone di ricerca destinato ad essere coltivato a lungo, con lo sviluppo, parallelo e indipendente, di due programmi in grado di giocare a dama contro avversari umani, rispettivamente da parte di Christopher Strachey dell'Università di Manchester e Arthur Samuel dell'IBM[75]. Una dimostrazione televisiva delle capacità del programma di Samuel, nel 1956, fece molto scalpore tra il pubblico americano[76].

Sulla base di questi primi successi, la conferenza di Dartmouth fu accompagnata da un clima di diffuso entusiasmo attorno alle possibilità dell'intelligenza artificiale. La stessa più volte citata *Proposal* della conferenza aveva obiettivi molto ambiziosi, ritenendo possibile raggiungere risultati significativi già in quella sede: «*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*»[77].

In realtà, l'evento di Dartmouth si concluse senza risultati rivoluzionari, risolvendosi in un'occasione di incontro e avvio di collaborazioni scientifiche[78], più che di continuativo lavoro di ricerca. Ciò nonostante, tale clima di fiducia permase per l'intero decennio successivo, caratterizzato da significativi progressi nel campo, che fecero pensare – come vedremo, erroneamente - che gran parte degli obiettivi enunciati nel 1956 fosse raggiungibile in tempi relativamente brevi.

2. *General Problem Solver*, Lisp, ELIZA e i primi successi

Il 1957 è l'anno in cui Newell e Simon misero a punto il già citato *General Problem Solver*, un sistema in grado di risolvere problemi di logica e geometria euclidea imitando il percorso del ragionamento umano[79].Il programma rappresentava il completamento di un lavoro presentato l'anno prima a Dartmouth con il nome di *LogicTheorist*[80]. È molto nota un'iperbolica dichiarazione di

[74]A. TURING, *Computing machinery and intelligence cit.*

[75] C.S. STRACHEY, *Logical or non-mathematical programs*, in *ACM '52: Proceedings of the 1952 ACM national meeting*, 1952, p. 46-49; A. SAMUEL, *Some studies in machine learning using the game of checkers*, in *IBM Journal*,3, 1959, p. 211-229; J. SCHAFFER, *Solving the Game of Checkers*, in *Mathematical Sciences Research Institute Publications*, 29, 1996, p. 119-131.

[76]S.J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (4ª ed.)*, p. 19.

[77] J. MCCARTHY, M.L. MINSKY, N. ROCHESTER, C.E. SHANNON, *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence cit.*, p. 2.

[78] Cfr. ancora S.J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (4ª ed.)*, p. 18.

[79]A. NEWELL, J. C. SHAW, H. A. SIMON, *Report on a general problem-solving program cit.*

[80]A. NEWELL, H. SIMON, *The logic theory machine: a complex information processing system*, RAND Corporation - report, 15 giugno 1956.

Simon dello stesso anno, adatta a restituire il clima di fiducia attorno alle potenzialità dell'intelligenza artificiale che si era creato in quel periodo:

«It is not my aim to surprise or shock you – but the simplest way I can summarize is to say that there are now in the world machines that think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until – in a visible future – the range of problems they can handle will be coextensive with the range to which the human mind has been applied»[81].

In realtà, i rapidi progressi nel campo sembrarono giustificare tale entusiasmo, almeno sino alla seconda parte degli anni '60. Nel 1958 John McCarthy, nel frattempo passato ai laboratori del MIT, creò il linguaggio di programmazione di alto livello Lisp, destinato a dominare la scena dell'intelligenza artificiale nei tre decenni successivi e all'origine di una famiglia di linguaggi usati ancora oggi[82].Nello stesso anno iniziò a lavorare nei laboratori del MIT anche Marvin Minsky, che vi spenderà l'intera carriera, fondando il menzionato *MIT Artificial Intelligence Project*[83](McCarthy, invece, si trasferirà a Stanford 5 anni dopo, dove avvierà l'*AI Lab*)[84]. Nel corso degli anni '60, Minsky coordinerà il lavoro di diversi studenti e giovani ricercatori del MIT[85], che concentreranno i loro sforzi nello sviluppo di programmi in grado di risolvere problemi in campi molto limitati della conoscenza, i c.d. "micromondi"[86]. Furono sviluppati algoritmi in grado di calcolare integrali simili a quelli proposti a studenti del *college*[87], risolvere quesiti di analogia geometrica come quelli inclusi nei test del QI[88], o organizzare la sistemazione di alcuni oggetti nello spazio[89].

Parallelamente, si registrarono progressi anche nel campo delle reti neurali. Sempre nel 1958, lo psicologo statunitense Frank Rosenblattpropose per la prima volta il percettrone, un modello base di rete neurale che sarà, decenni dopo, il punto di partenza per lo sviluppo di reti complesse[90]. Il

[81] La citazione è riportata, tra gli altri, da S.J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (4ª ed.)*, p. 21; D. CREVIER, *AI: The Tumultuos Search for Artificial Intelligence cit.*, p. 17.

[82]J. MCCARTHY, *Recursive functions of symbolic expressions and their computation by machine*, in *Communications of the ACM*, April 1960; D. HEMMENDINGER, *LISP – computer language*, in *Encyclopedia Britannica*, 2016, https://www.britannica.com/technology/LISP-computer-language (28 marzo 2021).

[83] Oggi uno dei centri di ricerca più rilevanti a livello globale, cfr: https://www.csail.mit.edu/ (28 marzo 2021).

[84] Si rimanda, anche in questo caso, alla pagina web della struttura: https://ai.stanford.edu/ (28 marzo 2021).

[85] Alcuni dei quali, in seguito, diedero un rilevante contributo allo sviluppo della disciplina, come gli informatici Robert Seagle, poi ricercatore al MIT e professore all'Università di Minneapolis, e Daniel G. Bobrow, poi sviluppatore del sistema operativo TENEX.

[86]Cfr. S.J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (4ª ed.)*, p. 20-21.

[87] J.R. SLAGE, *A heuristic program that solves symbolic integration problems in freshman calculus: symbolic automatic integrator (SAINT)*, 1961, https://dspace.mit.edu/handle/1721.1/11997 (28 marzo 2021).

[88] T.G. EVANS, *A program for the solution of geometric analogy intelligence test questions*, in M. MINSKY, *Semantic Information Processing*, Cambridge (US) 1968, p. 271-353.

[89] Il micromondo costituito da una superficie piana su cui muovere una serie di solidi geometrici (c.d. *blocks world*) è stato senza dubbio il più popolare, venendo usato nello sviluppo di molteplici programmi, v. J. SLANEY, *Blocks World revisited*, in *Artificial Intelligence*, 125, 1-2, 2001, p. 119-153.

[90] F. ROSENBLATT, *The Perceptron - a perceiving and recognizing automaton*, Report 85-460-1 - Cornell Aeronautical Laboratory, Buffalo (US), 1957, https://bit.ly/3cu7Zlj (28 marzo 2021).

percettrone di Rosenblatt era composto da un'unità d'ingresso, un'unità d'uscita e una regola di apprendimento basata sulla minimizzazione dell'errore (*error back-propagation*). Grazie alla *back-propagation*, il modello di Rosenblatt era in grado di adattare il peso numerico associato a ciascuna connessione fino a che l'output non fosse quello desiderato. Il percettrone, dunque, era capace di memorizzazione e apprendimento, un'eventualità solo ipotizzata da Pitts e McCullogh[91].

Infine, ottennero grande risonanza i primi risultati concreti in uno dei settori enumerati dalla proposta di Dartmouth, ma fino ad allora pressoché privo di riscontri pratici: l'elaborazione del linguaggio naturale. È rimasto celebre *ELIZA*, il primo chatbot della storia, sviluppato nei laboratori del MIT tra il 1964 e il 1966[92]. Concepito dal suo creatore Joseph Weizenbaum come la parodia di un terapeuta rogersiano[93], il programma era in grado di sostenere conversazioni relativamente semplici in lingua inglese, per mezzo di schemi di risposta predeterminati in cui riutilizzava le parole dell'interlocutore. Nonostante lo stesso Weizenbaum lo considerasse una dimostrazione di come, allo stato dell'arte, ogni tentativo di comunicazione linguistica uomo-macchina fosse limitato da alti livelli di artificiosità e superficialità, fu sorprendente notare quante persone credettero, in realtà, che il programma fosse un essere umano in carne ed ossa, anche dopo conversazioni di lunghezza considerevole[94].

3. Le battute d'arresto, il primo inverno dell'IA e i sistemi esperti

Le previsioni ottimistiche formulate nel decennio precedente si scontrarono con la realtà nei primi anni '60, quando i rapidi progressi che si attendevano nel campo dell'intelligenza artificiale tardarono a realizzarsi. La ragione principale fu la difficoltà di applicare le prime tecnologie sviluppate nel campo ai problemi del mondo reale[95]. Infatti, i programmi pensati per la gestione di micromondi funzionavano con la manipolazione di un numero molto basso di variabili e scenari possibili. La loro applicazione a problemi concreti si rilevò estremamente difficile a causa della necessità di interfacciarsi con una quantità esponenzialmente maggiore di oggetti e possibilità. Il numero di potenziali soluzioni tra cui individuare quella corretta diventava estremamente alto, richiedendo una complessità di calcolo del tutto fuori portata per i computer del tempo (si tratta del

[91] D. CREVIER, *AI: The Tumultuos Search for Artificial Intelligence cit.*, p. 120-125; S.J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (4ª ed.)*, p. 21.
[92] J. WEIZENBAUM, *ELIZA – A computer program for the study of natural language communication between man and machine*, in *Communications of the ACM*, 9, 1, 1966, http://www.universelle-automation.de/1966_Boston.pdf (22 marzo 2021).
[93] Presupposto fondamentale della psicoterapia rogersiana è l'instaurazione di una relazione paritaria tra terapeuta e cliente, cfr. C. ROGERS, *Client-Centered Therapy: Its Current Practice, Implications and Theory*, Londra, 1951.
[94] Cfr. D. CREVIER, *AI: The Tumultuos Search for Artificial Intelligence cit.*, p. 132-144; N.J. NILSSON, *The Quest for Artificial Intelligence*, p. 38-40. Una versione online di ELIZA è oggi disponibile a questo link: http://psych.fullerton.edu/mbirnbaum/psych101/Eliza.htm (22 marzo 2021).
[95] S.J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (4ª ed.)*, p. 21-22.

problema noto, in linguaggio matematico, come "esplosione combinatoria"[96]). L'intelligenza artificiale iniziò, così, a venire considerata una tecnologia capace di risolvere solo "*toy problems*", inadatta ad applicazioni utili a fini scientifici e industriali[97].

Contemporaneamente, si registrò una decisa battuta d'arresto nel campo delle reti neurali, anche con il contributo di uno dei "padri" dell'intelligenza artificiale, Marvin Minsky. Nel 1969 Minsky pubblicò insieme all'informatico e pedagogista Seymour Papert, anch'egli ricercatore nei laboratori del MIT, il volume *Perceptrons: an Introduction to Computational Geometry*[98], che ebbe una grande influenza sul dibattito scientifico del decennio successivo. Nel libro, i due studiosi dimostrarono alcuni limiti delle reti neurali basate sul modello del percettrone di Rosenblatt. Minsky e Papert resero evidente, in particolare, che il percettrone non poteva essere utilizzato per la soluzione di alcuni tipi di funzioni e che un percettrone a due input non poteva essere addestrato per riconoscere la differenza tra di essi. Tali limiti, in realtà, sarebbero stati poi superati con lo sviluppo di reti sufficientemente profonde, ma l'impatto dell'opera di Minsky e Papert fu tale da portare all'abbandono quasi totale della ricerca nel campo delle reti neurali almeno fino a metà degli anni '80[99].

Le difficoltà applicative portarono alla messa in discussione degli investimenti per la ricerca nel settore, fino ad allora molto ingenti. Il governo britannico, nel 1972, commissionò a James Lighthill, professore di matematica applicata all'università di Cambridge, il rapporto *Artificial intelligence: a general survey*[100], che espresse giudizi moderatamente pessimistici sulle possibilità di sviluppo a breve termine dell'intelligenza artificiale, ponendo l'accento sulla sottovalutazione del problema dell'esplosione combinatoria[101]. Il documento portò alla decisione di revocare i finanziamenti alla ricerca nel campo, meno che in due università del paese. Provvedimenti simili furono presi anche dal governo degli Stati Uniti, che nello stesso anno diminuì drasticamente gli investimenti messi in atto fino ad allora dal Ministero della difesa attraverso la Defence Advanced

[96]L. SHERREL, *Combinatorial explosion*, in RUNEHOV A.L.C., OVIEDO L. (A CURA DI), *Encyclopedia of Sciences and Religions*, Dordrecht, 2013, doi:10.1007/978-1-4020-8265-8_201037; K. KRIPPENDORFF, *Combinatorial Explosion,* in *Web Dictionary of Cybernetics and Systems*, http://pespmc1.vub.ac.be/ASC/COMBIN_EXPLO.html (28 marzo 2021).

[97]S.J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (4ª ed.)*, p. 21-22; N. J. NILSSON, *Artificial intelligence: a new synthesis cit.*, p. 8-9.

[98]M. MINSKY, S. PAPERT, *Perceptrons: an introduction to computational geometry*, Cambridge (US), 1969.

[99]H.D. BLOCK, *A review of "Perceptrons: an introduction to computational geometry"*, in *Information and control*, 17, 1970, p. 501-522; S.J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (4ª ed.)*, p. 22.

[100]J. LIGHTHILL, *Artificial intelligence: a general survey by Sir James Lighthill*, *FRS Lucasian Professor of Applied Mathematics*, Cambridge, 1972, https://aitopics.org/doc/classics:D8235CF9/ (28 marzo 2021).

[101]Il rapporto, infatti, additava l'esplosione combinatoria come la causa principale del mancato raggiungimento delle aspettative riposte nell'IA: «[…] the whole of this report represents in the last analysis only the personal view of the author. Before going into such detail he is inclined as a mathematician to single out one rather general cause for the disappointments that have been experienced: failure to recognise the implications of the "combinatorial explosion". This is a general obstacle to the construction of a self-organising system on a large knowledge base which results from the explosive growth of any combinatorial expression, representing numbers of possible ways of grouping elements of the knowledge base according to particular rules as the bases size increases.», *Ibidem*, p. 9.

Research Project Agency (DARPA)[102]. La situazione fece venir meno le risorse ai più importanti centri di ricerca impegnati nell'ambito e causò, a cascata, la chiusura o l'abbandono del settore da parte di diverse imprese private che vi si erano avventurate. Gli specialisti sono soliti riferirsi a questo periodo con l'espressione *primo inverno dell'intelligenza artificiale*[103] (ve ne sarà, come si dirà, un secondo) per enfatizzare la stagnazione di progressi e investimenti che lo caratterizzò.

L'interesse per l'intelligenza artificiale tornò a crescere solo con l'avvento dei c.d. sistemi esperti[104]. Come già detto, uno dei problemi essenziali riscontrati nell'applicazione pratica dell'intelligenza artificiale era stata la complessità e l'alto numero di variabili che caratterizzano il mondo reale. Alla base dei sistemi esperti c'era l'idea di ovviare al problema sviluppando strumenti in grado di operare in campi molto ristretti della realtà, applicando conoscenze e regole di decisione codificate con la programmazione. La concezione di fondo era simile a quella che aveva ispirato lo sviluppo dei micromondi, replicata su scala più larga grazie all'accresciuta potenza computazionale resasi disponibile con lo sviluppo tecnologico[105].

Tale intuizione portò alle prime applicazioni dell'intelligenza artificiale apprezzate dal mondo produttivo. Uno dei primi e più noti esempi è DENDRAL, un programma specializzato nella lettura di spettrometrie di massa il cui prototipo fu sviluppato nel 1969 nei laboratori di Stanford dall'allievo di Herbert Simon Ed Feigenbaum, col filosofo Bruce Buchanan e il genetista Joshua Lederberg[106]. Sempre a Stanford, nella prima metà degli anni '70 fu sviluppato MYCIN, un sistema esperto in grado di riconoscere infezioni batteriche e consigliare antibiotici, che, nonostante non sia mai stato utilizzato nella pratica, aprirà la strada alle numerose applicazioni successive dell'intelligenza artificiale in ambito medico[107]. I sistemi esperti conquistarono rapidamente crescenti porzioni di mercato negli anni '80. Il primo programma a ottenere un significativo successo commerciale fu R1, adottato dalla azienda informatica statunitense Digital Equipment Corporation nel 1980 per rendere più efficiente la distribuzione dei propri prodotti[108]. Si stima che il

[102] D. CREVIER, *AI: The Tumultuos Search for Artificial Intelligence*, p. 108 ss; S.J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (4ª ed.)*, p. 21-22.
[103] M. LIM, *History of AI winters*, in *Actuaries Digital*, 2018, https://bit.ly/2Pd8w2u (30 marzo 2021); S. SCHUCHMANN, *History of the first AI winter*, in *Towards data science*, 2019, https://bit.ly/3u6jbdV (30 marzo 2021).
[104] Cfr. tra I molti N.J. NILSSON, *The Quest for Artificial Intelligence cit.*, p. 229 ss.
[105] S.J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (4ª ed.)*, p. 22-24;D.CREVIER,*AI: The TumultuosSearch for Artificial Intelligence*, p. 146 ss.
[106] E.A. FEIGENBAUM, B. BUCHANAN, J. LEDERBERG, *Ongenerality and problem solving: a case study using the DENDRAL program*, Stanford Artificial Intelligence Project – Computer Science Dept. Report n. CS176, i.stanford.edu/pub/cstr/reports/cs/tr/70/176/CS-TR-70-176.pdf (28marzo2021); R. K. LINDSAY, B.G. BUCHANAN, E.A. FEIGENBAUM, J. LEDERBERG, *DENDRAL: a case study of the first expert system for scientific hypothesis formation*, in *Artificial Intelligence*, 61, 1993, p. 209-261.
[107] E. SHORTLIFFE, B.G. BUCHANAN, *A model of inexact reasoning in medicine*, in *Mathematical Biosciences*, 23, 3-4, 1975, p. 351-379; B.G. BUCHANAN, E. H. SHORTLIFFE, *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project*, Reading (USA), 1994; v. ancheN. J. NILSSON, *Artificial intelligence: a new synthesis*, Burlington, 1998, p. 229.
[108] J. MCDERMOTT, *R1: a rule-based configurer of computer systems*, in *Artificial Intelligence*, 19, 1, 1982, p. 39-88.

sistema abbia consentito alla compagnia di risparmiare circa 40 milioni di dollari l'anno[109]. Per la fine del decennio, pressoché ogni grande impresa americana utilizzava sistemi esperti. Il volume d'affari connesso all'intelligenza artificiale passò, in meno di un decennio, da pochi milioni a miliardi di dollari, con la nascita di decine di imprese specializzate[110]. Tale situazione ebbe ripercussioni, ovviamente, anche sul fronte dei finanziamenti alla ricerca, che tornarono a crescere.

4. La ripresa del settore, il ritorno delle reti neurali e il secondo inverno dell'IA

La spinta decisiva per la ripresa del finanziamento pubblico alla ricerca sull'intelligenza artificiale venne dal Giappone. Nel 1981, infatti, la monarchia orientale annunciò il programma di investimenti decennale *Fifth Generation*[111], invitando il Regno Unito a prendervi parte come partner. Il governo britannico, come dieci anni prima, commissionò a un gruppo di esperti un'indagine sulle potenzialità di sviluppo del settore, incaricando l'allora direttore tecnico di British Telecom John Alvey[112]. Le valutazioni positive contenute nel rapporto di Alvey convinsero il governo a rifiutare la proposta di partnership del Giappone e avviare un autonomo piano nazionale di investimenti nel settore tecnologico, noto, appunto, come *Alvey Programme*[113]. Parallelamente, negli Stati Uniti, per tenere il passo di Giappone e Gran Bretagna,fu creato il Microelectronics and Computer Consortium, che riuniva alcune delle più rilevanti imprese operanti nel campo e i centri di ricerca pubblici e privati più avanzati. Anche la già menzionata agenzia pubblica DARPA, tra il 1984 e il 1988, triplicò gli investimenti[114].

Oltre a vedere il rifiorire dei finanziamenti al settore, gli anni '80 furono anche il momento del ritorno delle reti neurali. Pressoché contemporaneamente, l'algoritmo di retropropagazione per l'addestramento delle reti neurali fu "riscoperto" e reso popolare da diversi gruppi di ricerca[115]. Nel 1982, inoltre, il fisico americano John Hopfield presentò un modello di rete neurale, da allora noto, appunto, come rete di Hopfield, in grado di generare memoria c.d. associativa, utilizzando le

[109]Il dato è riportato da S.J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (4ª ed.)*, p. 23.
[110]*Ibidem.*
[111]E. H. SHAPIRO, *The fifthgeneration project – a trip report*, in *Communications of the ACM*, 26, 9, 1983, p. 637-641; M. CROSS, *Japan's fifth generation computer project successes and failures*, in *Futures*, 21, 4, 1989, p. 401-403; E. FEIGENBAUM, P. MCCORDUCK, *The fifth generation: artificial intelligence and Japan's computer challenge to the world*, Boston, 1983.
[112]A.A.V.V., *A programme for advanced information technology. The Report of the Alvey Committee*, Londra, 1982, https://stacks.stanford.edu/file/druid:wg645hn3953/wg645hn3953.pdf (23 marzo 2021).
[113]D. B. THOMAS, *The Alveyprogramme – intelligent knowledge-based systems aspects*, in *R&D Management*, 15, 2, 1985, p. 101-103.
[114]D. CREVIER,*AI: The Tumultuos Search for Artificial Intelligence*, p. 197 ss.; S.J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (4ª ed.)*, p. 22-24; N.J. NILSSON, *The Quest for Artificial Intelligence cit.*, p. 286.
[115] Si vedanoadesempio i lavori, condottipressoché in contemporanea, di D.E. RUMELHART, G.E. HINTON, R.J. WILLIAMS, *Learning representations by back-propagating errors*, in *Nature*, 323, 6088, 1986, p. 533–536; D.B. PARKER, *Learning Logic*, MIT Center for Computational Research in Economics and Management Science – Technical Report, 1985; Y. LECUN, *Une procedure d'apprentissage pour réseau a seuil asymmetrique (a Learning Scheme for Asymmetric Threshold Networks)*, in *Proceedings of Cognitiva 85*, Paris,1985, p. 599-604.

conoscenze accumulate con l'esperienza per operare in contesti di informazione scarsa o deteriorata (come nel riconoscimento di immagini parzialmente incomplete)[116]. I progressi più significativi nel campo furono raccolti in un volume collettaneo del 1986, *Parallel Distributed Processing*, a cura degli psicologi statunitensi James McClelland e David Rumelhart[117], il cui notevole impatto scientifico contribuì ad affermare definitivamente le reti neurali come uno dei filoni più promettenti dell'intelligenza artificiale.

Il volume d'affari crebbe incontrollato fino alla fine degli anni '80, quando il settore registrò alcuni segni di stagnazione, che preoccuparonoi ricercatori che avevano assistito, nel decennio precedente, al primo crollo dei finanziamenti[118]. In particolare, nel 1987 i personal computer di Apple e IBM superarono in rendimento le macchine programmate col menzionato linguaggio Lisp, allora dominante sulla scena dell'intelligenza artificiale. Nello stesso periodo, la diffusione sul mercato dei sistemi esperti iniziò a rallentare a causa della difficoltà di aggiornarli per il loro impiego in contesti di crescente complessità, anche alla luce del fatto che, al contrario delle reti neurali, erano incapaci di autoapprendimento[119]. Più in generale, ancora una volta le aspettative sull'intelligenza artificiale avevano superato le reali possibilità di progresso: basti pensare che alcuni degli obiettivi che il piano *Fifth generation* del Giappone aveva fissato per il 1991 non possono dirsi pienamente raggiunti nemmeno ai giorni nostri, come lo sviluppo di macchine capaci di «*carry on a casual conversation*»[120]. Tra il finire degli anni '80 e l'inizio degli anni '90, così, i fondi per la ricerca furono drasticamente ridotti da tutte le principali agenzie pubbliche, e migliaia di imprese che operavano nel campo dell'intelligenza artificiale furono costrette a chiudere o convertire la loro attività[121]. Generalmente, questo periodo è chiamato "secondo inverno dell'intelligenza artificiale", per evidenziare il nuovo rallentamento nello sviluppo tecnologico che lo caratterizzò[122]. La ripresa che vi seguirà, a partire da metà degli anni '90, proseguirà ininterrotta sino ai giorni nostri.

---

[116] 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, p. 2554–2558; L. MACERA, *Memorie associative e reti di Hopfield*, in *MCmicrocomputer*, 105, 1991, p. 282-285.

[117] D.E. RUMELHART, J. L. MCCLELLAND (A CURA DI), *Parallel distributed processing*, Cambridge (US), 1986

[118] La crisi che investì il settore fu anticipata, ad esempio, in un dibattito al congresso dell'Association for the Advancement of Artificial Intelligence del 1984, dal titolo eloquente *the Dark Ages of AI*, cfr. D. MCDERMOTT, M. MITCHELL WALDROP, R. SCHANK, B. CHANDRASEKARAN, J. MCDERMOTT, *The dark ages of AI: a panel discussionat AAAI-84*, in *AI Magazine*, 6, 3, 1985, p. 122-134.

[119] D. CREVIER, *AI: The Tumultuos Search for Artificial Intelligence cit.*, p. 209-210; P. MCCORDUCK, *Machines Who Think (2ª ed.)*, Natick, 2004, p. 435.

[120] Cfr. P. MCCORDUCK, *Machines Who Think cit.*, p. 441.

[121] H.P. NEWQUIST, *The brain makers: genius, ego, and greed in the quest for machines that think*, Indianapolis, 1994, p. 380; v. ancheH. MORAVEC, *The great 1980s AI bubble: a review of the brain makers*, in *AI Magazine*, 15, 3, 1994, p. 86-87.

[122] V. ancora M. LIM, *History of AI winters cit.*, https://bit.ly/2Pd8w2u (30 marzo 2021); S. SCHUCHMANN, *History of the first AI winter cit.*, https://bit.ly/3u6jbdV (30 marzo 2021).

5.Gli anni ’90 e ’00: l’impiego su larga scala dell’intelligenza artificiale, la “*victory of the neats*” e i *big data*

Il ritorno di ingenti investimenti nel settore coincise con la crescita della capacità computazionale dei computer, che permise lo sviluppo di algoritmi sempre più complessi e utili per i processi produttivi, senza scontrarsi con l’ostacolo, già menzionato, dell’esplosione combinatoria[123].La seconda metà degli anni ’90 vide la definitiva affermazione dell’intelligenza artificiale in diversi ambiti chiave del mondo economico, spesso all’interno di sistemi tecnologici e organizzativi più ampi, non composti unicamente da strumenti intelligenti. L’IA acquistò, così, il ruolo fondamentale in campi come la logistica, la gestione del sistema bancario, o certi filoni della diagnostica medica che riveste ancora oggi, e furono sviluppate nuove tecnologie e applicazioni software destinate a diventare insostituibili, come i motori di ricerca, i sistemi di data mining e traduzione automatica, o la moderna robotica industriale[124]. Nello stesso periodo, l’11 maggio del 1997, avvenne uno degli eventi simbolo per lo sviluppo dell’intelligenza artificiale, il primo capace di attirare l’attenzione del grande pubblico: la vittoria di *Deep Blue*, un computer sviluppato da IBM per il gioco degli scacchi, contro il campione del mondo dell’epoca Garry Kasparov[125].

Dal punto di vista teorico e tecnico, i progressi avvenuti tra la seconda metà degli anni ’90 e la prima degli anni 2000 mossero, in generale, verso l’adozione di un approccio via via più scientifico e aperto al contributo di altre discipline. Divenne più comune, rispetto ai primi decenni di storia dell’intelligenza artificiale, cercare di migliorare teorie già formulate e tecnologie già presenti, invece di proporre tesi radicalmente nuove. La teoria della probabilità, gli studi sul linguaggio e i risultati ottenuti dalle scienze matematiche, statistiche ed economiche riguardo ai procedimenti di decisione e di scelta razionale contaminarono in modo crescente il settore[126]. Questa evoluzione è

[123] S.H. FULLER, L.I. MILLET, *The future of computing performance. Game over or nextlevel?*, Washington, 2011, p. 155; per una cronologia della crescita della capacità computazionale dei computer negli ultimi decenni, v. M. ROSER, H. RITCHIE, *Technological progress*, in *Our world in data*, 2013, https://ourworldindata.org/technological-progress (31 marzo 2021).

[124] Su queste applicazioni dell’IA si vedano, ad esempio, S. J. RUSSELL, P. NORVIG, *Artificial Intelligence cit. (3ª ed.)*, p. 24-27; W.J. HUTCHINS, *Machine translation: a brief history*, in E.F.K. KOERNER, R.E. ASHER (a cura di),*Concise history of the language sciences. From the Sumerians to the cognitivists*, Amsterdam, 1995, p. 431-445; V. KAUL, S. ESLIN, S.A. GROSS, *History of artificial intelligence in medicine*, in *Gastrointestinal Endoscopy*, 92, 4, 2020, p. 807-812; S.OLSEN, *Spying an intelligent search engine*, in *www.cnet.com*, 2006, https;//cnet.co/2Z4ccVN (31 marzo 2021).

[125] B. WEBER, *Swift and slashing, computer topples Kasparov*, The New York Times, 12 maggio 1997; la storia dello sviluppo di Deep Blue è consultabile sul sito della stessa IBM, https://ibm.co/2SFhrqC(29 marzo 2021).

[126] Stuart Russell e Peter Norvig, nella terza edizione del loro celebre manuale *Artificial intelligence: a modern approach*, a p. 25, si riferivano al fenomeno con l’efficace espressione *«AI adopts the scientific method»*. Per diversiesempidell’adozione di un approccioscientifico-sperimentalenell’approccioall’intelligenza artificial neglianni ’90 v. P.R. COHEN, *Empirical methods for artificial intelligence*, Cambridge (US), 1995; D.A. MCALLESTER, *What is the most pressing issue facing AI and the AAAI today?*,Candidate statement, election for Councilor of the American Association for Artificial Intelligence, 1998, chedescriveva la crescente contaminazione del settore con le parole: «In the early period of AI it seemed plausible that new forms of symbolic computation,e.g., frames and semantic networks, made much of classical theory obsolete. This led toa form of isolationism in which AI became largely separated from the rest of computerscience. This isolationism is currently being abandoned. There is a recognition thatmachine learning should not be isolated from information theory, that uncertain reasoningshould not be isolated from stochastic

stata definita "*victory of the neats*", ricostruendo le vicende storiche dell'intelligenza artificiale come una contrapposizione tra *scruffies*, animati da un approccio puramente sperimentale, per cui l'IA si sarebbe sviluppata tentando idee sempre diverse e selezionando le migliori, e *neats*, convinti che lo sviluppo dell'IA dovesse procedere con rigore matematico e teorico, la cui visione avrebbe finito per prevalere. In realtà, la successiva diffusione del *deep learning* potrebbe essere interpretata come un ritorno della visione *scruffy*[127].

È in questo periodo che si affermò definitivamente l'approccio all'IA basato sul concetto di agente razionale, il cui sviluppo divenne l'obiettivo condiviso di gran parte degli studiosi, generando un'inedita condivisione di risultati e strategie. Diventò abituale, in particolare, la messa in comune di *dataset* su cui lavorare e con cui misurare i progressi ottenuti nei vari campi, come le moderne banche dati LibriSpeech per il riconoscimento del linguaggio naturale[128], o ImageNet per quello delle immagini[129]. Questa crescente disponibilità di dati, assieme alla già menzionata maggiore capacità computazionale delle macchine, portò a rapidi progressi nel campo dell'apprendimento automatico, le cui applicazioni pratiche crebbero e si diversificarono molto, e che in quel periodo divenne comune chiamare, anche in lingua italiana, *machine learning*.

Si trattava, in realtà, solamente del preludio di quanto sarebbe avvenuto nel primo decennio degli anni 2000, quando la diffusione del world wide web portò a un aumento esponenziale del volume di dati a disposizione e a un'inedita capacità di immagazzinarli, catalogarli e condividerli online (c.d. *big data*)[130]. Le possibilità dischiuse dai *big data* per lo sviluppo dell'intelligenza artificiale superarono ogni aspettativa, con strabilianti progressi, ad esempio, nei già nominati campi del riconoscimento delle immagini e dell'elaborazione del linguaggio naturale[131]. Risultò in breve tempo evidente, in particolare, che aumentando la grandezza dei *dataset* su cui allenare gli algoritmi era possibile ottenere risultati fino ad allora mai ottenuti, in molti casi assimilabili alle performance di un agente umano. I due sviluppatori di Microsoft Michele Banko ed Erik Brill, nel 2001, notarono che i progressi che era possibile ottenere semplicemente allargando i *dataset* di partenza superavano di gran lunga quelli a cui si poteva ambire raffinando la struttura degli algoritmi,

---

modeling, that search should not be isolated fromclassical optimization and control, and that automated reasoning should not be isolatedfrom formal methods and static analysis».

[127]Per la contrapposizione tra *neats* e *scruffies*, e la provvisoria vittoria dei primi, v. P. MCCORDUCK, *Machines Who Think cit.*, p. 487; S. RUSSELL, P. NORVIG, *Artificial intelligence cit. (4ª ed.)*, p. 24.

[128] Un database contenente più di 1000 ore di registrazioni di audiolibri, cfr. V. PANAYOTOV, G. CHEN, D. POVEY, S. KHUDANPUR, *LibriSpeech: an ASR corpus based on public domain audiobooks*, in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, 2015, p 5206-5210, doi:10.1109/ICASSP.2015.7178964.

[129] Un database di più di 14 milioni di immagini divise in categorie e annotate con l'elenco degli oggetti che vi compaiono,http://image-net.org/index (29 marzo 2021).

[130]Tra I molti, v. C. SNJIDERS, U. MATZAT, U. D. REIPS, *"Big Data": Big Gaps of Knowledge in the Field of Internet Science*, in *International Journal of Internet Science*, 7, 2012, p, 1-5.

[131]Per una panoramica su tali risultati si veda ancora S. RUSSELL, P. NORVIG, *Artificial intelligence cit.(4ª ed.)*, p. 26 ss.

dimostrando come disporre di grandi moli di dati fosse prioritario rispetto alla stessa qualità tecnica del sistema chiamato ad elaborarli[132].

Anche grazie alla rivoluzione rappresentata dai *big data*, all'inizio degli anni 2000 si registrarono due avvenimenti dall'alto valore simbolico, che contribuirono a consolidare l'interesse pubblico e mediatico per gli avanzamenti dell'intelligenza artificiale, quasi un decennio dopo la vittoria di *Deep Blue* contro Kasparov. Nel 2005 un'auto a guida autonoma completò per la prima volta il percorso, lungo 212 km, della DARPA Grand Challenge, una gara riservata a prototipi di auto senza conducente, rendendo evidenti le potenzialità di queste ultime[133]. Nel febbraio 2011, invece, Watson, un computer messo a punto da IBM con tecniche di *machine learning*, rappresentazione della conoscenza ed elaborazione del linguaggio naturale, sconfissei campioni del seguito quiz televisivo statunitense *Jeopardy!*, suscitando impressione nel pubblico e innescando una lunga discussione sulla stampa[134].

6. **L'avvento del *Deep Learning*, lo stato dell'arte e le prospettive future**

Successivamente alla loro "rinascita", negli anni '80,il progresso delle reti neurali è proseguito ininterrotto, con lo sviluppo di tipologie di reti estremamente diversificate per applicazioni e architettura, rallentando solamente durante la parentesi del menzionato "secondo inverno dell'IA"[135]. L'affermarsi di un approccio più condiviso e aperto alla contaminazione con altre discipline portò innovazioni anche in questo campo: in particolare, una pubblicazione di Judea Pearl del 1988, *Probabilistic Reasoning in Intelligent Systems*, aprì la strada all'implementazione della teoria della probabilità nelle reti neurali, permettendo lo sviluppo di algoritmi in grado di rappresentare in modo efficiente concetti non certi e formulare e affinare attraverso l'esperienza valutazioni probabilistiche[136].

Fu l'esplosione dei *big data*, però, a dare impulso a una vera e propria rivoluzione del settore, mettendo a disposizione *dataset* pressoché infiniti con cui "allenare le reti", in un mondo in cui la capacità computazionale non era più limitata come nei decenni precedenti. L'inedita potenza degli elaboratori permise lo sviluppo del *deep learning*, un composito insieme di tecniche di *machine learning* basato su reti neurali molto complesse, formate da diversi strati di neuroni artificiali. Nelle linee essenziali, ciò che caratterizza il *deep learning* è che il primo strato di neuroni della rete riceve

[132]M. BANCO, E. BRILL, *Scaling to Very Very Large Corpora for Natural Language Disambiguation*, in *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, 2001, p. 26-33, doi: 10.3115/1073012.1073017.
[133]J. MARKOFF, *In a grueling desert race, a winner, but not a driver*, The New York Times, 9 ottobre 2005.
[134]Vedi, ad esempio, J. MARKOFF, *Computer wins on "Jeopardy!": trivial' It's not*, The New York Times, 17 febbraio 2011; A. GABBAT, *IBM computer Watson wins Jeopardy clash*, The Guardian, 17 febbraio 2011.
[135]S. RUSSELL, P. NORVIG, *Artificial intelligence cit. (4ªed.)*, p. 24 ss.
[136]J. PEARL, *Probabilistic reasoning in intelligent systems*, Burlington, 1986.

l'input dall'esterno, mentre gli strati successivi utilizzano l'output dello strato precedente. Più la rete è profonda, dunque, più l'informazione è elaborata in modo completo e più l'output del sistema sarà accurato[137]. L'apprendimento avviene attraverso la memorizzazione e la modifica dei valori associati a un determinato input da ciascun nodo della rete, in analogia a quanto accade, secondo le attuali conoscenze neurobiologiche, coi legami sinaptici tra i neuroni del cervello umano[138].

Il *deep learning* è risultato particolarmente adatto all'analisi di grandi *dataset*. Presenta un'efficacia senza precedenti in operazioni di classificazione dei dati seguendo il modello di *training data* costituiti da esempi svolti (c.d. apprendimento supervisionato) e in attività di ricostruzione di categorie, correlazioni, modelli e inferenze senza esempi di partenza, su dati non organizzati (c.d. apprendimento non supervisionato)[139]. Si è imposto come una delle tecnologie dominanti il settore dell'intelligenza artificiale al principio degli anni '10 del 2000, quando le prime reti neurali profonde hanno portato a risultati fino a quel momento impensabili, inizialmente nel campo del riconoscimento delle immagini e del linguaggio[140]. Nel corso del decennio ne sono state sviluppate svariate applicazioni, portando alla diffusione commerciale su vasta scala di sistemi di riconoscimento vocale, traduzione automatica, diagnostica per immagini o decisione automatica avvenuta negli ultimi anni[141]. Questa evoluzione è andata di pari passo con i progressi tecnici nel campo, favoriti da investimenti pubblici e privati senza precedenti, che hanno reso trattabile con le reti neurali un numero sempre maggiore di problemi. Deve menzionarsi, per la rilevanza degli sviluppi tecnologici che ne sono derivati, il perfezionamento, tra il 2015 e il 2016, delle reti neurali *Long Short Term Memory*, in grado di analizzare non solo dati puntuali, ma anche la variazione di un insieme di dati in un intervallo temporale[142].

Come già avvenuto in altri momenti della storia dell'intelligenza artificiale, l'avvento del *deep learning* è stato accompagnato dalla vittoria di un sistema informatico contro un campione umano di un gioco di strategia. Tra il 9 e il 15 marzo del 2016, AlphaGo, un software sviluppato da

[137]Cfr. in generaleC.C. AGGARWAL, *Neural networks and deep learning – a textbook*, Berlino, 2018,p. 1441-1461; L. DENG, D. YU, *Deep Learning: Methods and Applications*, in *Foundations and Trends in Signal Processing*, 7, 3-4, p. 198-205, doi:10.1561/2000000039.

[138]Cfr. traimolti E. R. KANDEL, J. H. SCHWARTZ, T. M. JESSELL, S. A. SIEGELBAUM, A. J. HUDSPETH, *Principles of neural science (V^th ed.)*, New York, 2012, p. 1141-1161; J. L. MCCLELLAND, M. BOTVINICK, *Deep learning: Implications for human learning and memory,* in *The Oxford Handbook of Human Memory*, 2020.

[139]N.J. NILSSON, *The Quest for Artificial Intelligence cit.*, p. 413 ss.

[140]V. ad esempioA. KRIZHEVSKY, I. SUTSKEVER, G.E. HINTON, *ImageNet classification with deep convolutional neural networks*, in *NIPS*, 1, 2012, http://www.cs.toronto.edu/~fritz/absps/imagenet.pdf (30marzo2021).

[141]S. RUSSELL, P. NORVIG, *Artificial intelligence cit. (4ª ed.)*, p. 26-27.

[142] L'architettura *Long Short Term Memory* è stata teorizzata per la prima volta nel 1997 e perfezionata nel corso dei due decenni successivi. Tra il 2015 e il 2016 ha visto la prima applicazione pratica su larga scala, venendo implementata da Amazon, Google ed Apple nei rispettivi assistenti vocali e sistemi di traduzione automatica. Cfr. S. HOCHREITER, J. SCHMIDHUBER, *Long short-termmemory*, in *Neural Computation*, 9, 8, 1997, p. 1735-1780, doi: 10.1162/neco.1997.9.8.1735; T. CAPES, P. COLES, A. CONKIE, L. GOLIPOUR, A. HADJITARKHANI, Q. HU, N. HUDDLESTON, M. HUNT, J. LI, M. NEERACHER, K. PRAHALLAD, T. RAITIO, R. RASIPURAM, G. TOWNSEND, B. WILLIAMSON, D. WINARSKY, Z. WU, H. ZHANG, *Siri On-Device Deep Learning-Guided Unit Selection Text-to-Speech System*, in *Proceedings Interspeech*, 2017, p. 4011-4015, doi: 10.21437/Interspeech.2017-1798, p. 4011-4015.

Google, ha sconfitto a Go, uno dei giochi più complicati al mondo, tanto da prevedere $2{,}08\times10^{170}$ posizioni possibili delle pedine sulla plancia, il campione sudcoreano Lee Sedol[143]. L'attenzione mediatica riservata all'evento è uno dei fattori che ha contribuito a consolidare, nel dibattito generalista, il tema dell'intelligenza artificiale. La vittoria di AlphaGo, peraltro, non è rimasta isolata: tre anni dopo il software Pluribus ha battuto a poker texano quindici giocatori professionisti, un accadimento di particolare interesse vista la complessità del gioco, che rende necessarie abilità di negoziazione e in cui i criteri per definire sconfitta e vittoria non sono netti come negli scacchi o nel Go[144]. Negli stessi anni, sistemi intelligenti si sono affermati sui migliori giocatori umani in diversi videogiochi, come Dota II, StarCraft II e Doom, causando, in quest'ultimo caso, notevoli polemiche, posto che il gioco simula un contesto post-apocalittico in cui il protagonista sopravvive uccidendo i nemici con diverse armi da fuoco[145].

Allo stato dell'arte, molte delle potenzialità dell'intelligenza artificiale sembrano ancora agli inizi. Nonostante diverse previsioni formulate nei decenni passati si siano rivelate troppo avveniristiche, i risultati ottenuti, agli inizi del terzo decennio del XXI secolo, sono indubbiamente notevoli: basti pensare, oltre ai già citati progressi nei campi della robotica industriale, delle auto a guida autonoma, dell'elaborazione del linguaggio naturale o della diagnostica medica,alle applicazioni in campi come la geolocalizzazione o l'indicizzazione e la raccomandazione di contenuti online[146]. Le inedite capacità di analisi dei dati garantite dal *machine learning*, inoltre, sembrano offrire l'opportunità di elaborare modelli retrospettivi e predittivi molto complessi, che potrebbero dare un contributo chiave per affrontare alcune delle sfide della contemporaneità, a cominciare dal riscaldamento globale[147].

Contemporaneamente, l'intelligenza artificiale è al centro del dibattito scientifico e mediatico. Tra il 2010 e il 2019 il numero degli articoli scientifici nell'ambito è aumentato di venti volte, il numero degli studenti di corsi ad essa connessi nelle università è più che decuplicato, al pari di eventi e

[143] Si vedano, tra i molti, C. SAN-HUN, *Google's computer program beats Le-Sedol in Go tournament*, The New York Times, 16 marzo 2016, o, sulla stampa italiana, *AlphaGo ha vinto: la macchina ha battuto l'uomo 4-1*, La Repubblica, 15 marzo 2016. Storia e caratteristiche di AlphaGo possono consultarsi alla pagina web di DeepMind: https://deepmind.com/research/case-studies/alphago-the-story-so-far (29 marzo 2021).

[144] I. SAMPLE, *My poker face: AI wins multiplayer game for first time*, The Guardian, 11 luglio 2019; N. BROWN, T. SANDHOLM, *Superhuman AI for multoplayer poker*, in *Science*, 365, 6456, p. 885-890.

[145]N. STATT, *OpenAI'sDota 2 AI steamrolls world champion e-sports team with back-to-back victories*, The Verge, 13 aprile 2019, https://bit.ly/3u7dQD6; N. STATT, *DeepMind's StarCraft 2 AI is now better than 99.8 percent of all human players*, The Verge, 30 ottobre 2019; P. DOCKRILL, *Controversial AI Has Been Trained to Kill Humans in a Doom Deathmatch*, Science Alert, 1 ottobre 2016.

[146]Cfr. T. LI, Y. CUI, S. BELONGIE, J. HAYS, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, p. 5007-5015; A.SINGHAL, P.SINHA, R.PANT, *Use of Deep Learning in Modern Recommendation System: A Summary of Recent Works*, in *International Journal of Computer Applications*, 180, 7, 2017, p. 17-22.

[147] Una raccolta di più di 60 applicazioni dell'IA nel contrasto al riscaldamento globale può rinvenirsi in D. ROLNICK, P.L. DONTI, L.H. KAACK, K. KOCHANSKI, A. LACOSTE, K. SANKARAN ET AL., *TacklingClimate Change with Machine Learning*, in *ACM Computing Surveys*, 55, 2, 42, 2019.

conferenze nel campo, e sono comparse migliaia di aziende pronte a investire nel settore[148]. L'attenzione da parte della stampa, digitale e cartacea, è ormai costante e non più limitata a simboliche vittorie dei computer in determinati giochi di strategia, e i progressi nel campo sono sempre più noti anche al pubblico dei non esperti. Nella discussione, oltre ai giustificati entusiasmi, non mancano voci critiche e preoccupazioni. Come si dirà, infatti, alcune delle applicazioni dell'intelligenza artificiale presentano risvolti spinosi da vari punti di vista e pongono sfide inedite all'etica e al diritto[149]. Peraltro, va evidenziato che il dibattito etico, filosofico e antropologico sul rapporto tra essere umano e tecnologia precede l'avvento dell'intelligenza artificiale, presentando radici sorprendentemente risalenti nel tempo. L'attuale rinnovato interesse ne è semplicemente la naturale prosecuzione.

[148] Sono dati riportati da S. RUSSELL, P. NORVIG, *Artificial intelligence cit. (4 ed.)*, p. 27.

[149] Sul punto, è significativo che Stuart Russell e Peter Norvig, nella quarta edizione del celebre e più volte citato manuale *Artificial Intelligence – a modern approach*, risalente al 2020, abbiano introdotto, per la prima volta, un paragrafo dal titolo *Risks and benefits of AI* (p. 31 ss.).

L'etica delle macchine: brevi cenni sull'intelligenza artificiale nel pensiero filosofico

1. L'idea di macchine intelligenti e di meccanizzazione del pensiero: ipotesi e suggestioni dal mondo classico all'età moderna

L'idea di artefatti costruiti dall'uomo e dotati di autonomia è presente nella mitologia, nelle arti e nel pensiero filosofico fin dall'antichità, con almeno due esempi significativi già nel mito greco. In primo luogo, la vicenda, ripresa da Ovidio nelle *Metamorfosi*, dello scultore cipriota Pigmalione, innamoratosi perdutamente di una sua statua che la dea Afrodite, esaudendone le preghiere, trasformò in una donna in carne ed ossa[150]. In secondo luogo, il mito di Talos, il gigantesco automa di bronzo forgiato per Zeus da Efesto, dio del fuoco e della metallurgia, e posto dal re Minosse a guardia di Creta[151]. Riferimenti, inoltre, si rivengono anche nell'epica omerica. Nel XIV canto dell'Iliade la ninfa Teti, madre di Achille, si reca dallo stesso Efesto per chiedergli di forgiare delle nuove armi per il figlio e trova il dio circondato di assistenti robotici: ancelle modellate in oro, dotate di parola, e piccole strutture semoventi con tre piedi, dette, appunto, tripodi[152]. In ogni caso, creature artificiali dotate di vita propria sono presenti anche in altre tradizioni mitologiche del mondo antico: nella cultura ebraica, ad esempio, riveste importanza centrale la figura del golem, un gigante antropomorfo fabbricato a partire dall'argilla[153].

Nella storia del pensiero filosofico, è Aristotele il primo a riflettere con compiutezza sulla formalizzazione logica del pensiero razionale, che sarà, come visto nel capitolo precedente, una delle basi teoriche dello sviluppo dell'intelligenza artificiale. I suoi studi sul sillogismo, infatti, rappresentano il primo tentativo di codificare un sistema di regole con cui descrivere il ragionamento logico, a prescindere dai contenuti di volta in volta interessati[154]. In epoca moderna, l'idea che il ragionamento fosse replicabile attraverso la manipolazione di simboli è al centro del pensiero del filosofo, matematico e logico tedesco Gottfried Wilhem Leibniz (1646-1716), il primo

[150]PUBLIO OVIDIO NASONE, *Le Metamorfosi*, X, 243-297; riporta il mito anche il retore cristiano Arnobio (III sec. d.C.) in *Adversus nationes*, VI, 22.

[151]Narrano il mito PSEUDO-APOLLODORO, *Bibliotheca*, I, 9, 26; APOLLONIO RODIO, *Argonautiche*, IV, 1638-1693.

[152]OMERO, *Iliade*, XVIII, 505-580, nella traduzione di Vincenzo Monti (I[a] ed. Milano, 1810) le ancelle sono descritte «tutte d'oro e a vive/ giovinette simíli, entro il cui seno/ avea messo il gran fabbro e voce e vita/ evigor d'intelletto e delle care/ arti insegnate dai Celesti il senno», mentre i tripodi «Tutto in sudortrovollo affaccendato/ dÈ mantici al lavoro. Avea per mano/ dieci tripodi e dieci, adornamento/ di palagio regal. Sopposte a tutti/ d'oro avea le rotelle, onde ne gisse/ da sè ciascuno all'assemblea dÈ numi/ e da sè ne tornasse onde si tolse:/ maraviglia a vederli! Omai compiuto/ l'ammirando lavor, solo restava/ ch'ei v'adattasse le polite orecchie/ e appunto all'uopo n'aguzzava i chiovi.».

[153]Si rimanda allo studio di B. HENRY, *Dal golem ai cyborg. Trasmigrazioni nell'immaginario*, Livorno, 2013.

[154]Aristotele sviluppa la teoria del sillogismo negli *Analitici primi*, trattato che Andronico di Rodi ha poi raccolto nell'*Organon*, l'edizione delle sei opere di logica dello Stagirita, assieme alle *Categorie*, al *De Interpretazione*, agli *Analitici secondi*, ai *Topici* e agli *Elenchi sofistici*. Cfr. ARISTOTELE, *Organon*, a cura di G. COLLI, Torino, 1955.

a teorizzare la meccanizzazione del pensiero razionale. Leibniz sosteneva fosse possibile esprimere ogni concetto combinando un numero ristretto di proposizioni primitive, dette *characteristica universalis* [155]. Secondo la tesi di Leibniz, la *characteristica universalis* era la base per automatizzare il ragionamento: applicando un procedimento infallibile, il *calculus ratiocinator*, consistente nella combinazione delle proposizioni primitive coi principali operatori logici e che anche una macchina, appositamente costruita, avrebbe potuto eseguire, sarebbe stato possibile stabilire se un'affermazione fosse vera o falsa e risolvere quesiti di vario genere[156]. Liebniz teorizzava, addirittura, che i filosofi avrebbero potuto risolvere le loro dispute con tale marchingegno, semplicemente sottoponendovi le questioni su cui si trovavano in disaccordo[157]. Chiaramente, le teorie di Liebniz si scontrarono con la difficoltà di individuare quali dovessero essere le proposizioni primitive per formare l'universale *lingua characteristica*[158]. Ciò nondimeno, il pensatore tedesco fu il primo a teorizzare con chiarezza che attraverso la manipolazione di un numero finito di simboli fosse possibile automatizzare almeno parte del ragionamento razionale, ponendo un altro tassello fondamentale per la genesi dell'idea di intelligenza artificiale[159].

Le intuizioni di Liebniz furono sviluppate tra il XIX e il XX secolo, con la comparsa dei primi sistemi completi di simboli logici, con cui rappresentare graficamente enunciati e giudizi attraverso le relazioni tra le proposizioni che li compongono. Generalmente, il primo linguaggio formale è considerato l'*Ideografia* del filosofo e matematico tedesco Friedrich Ludwig Gottlob Frege[160] (1848-1925), mentre la notazione logica tuttora in uso sarà codificata, nel secolo successivo, prima dal matematico italiano Giuseppe Peano e poi dai britannici Bernard Russell e Alfred North Whitehead. Nel XIX secolo, inoltre, il logico George Boole teorizzò, in due pubblicazioni a pochi anni di distanza, *The Mathematical Analysis of Logic* (1847)[161] e *An Investigation of the Laws of Thought* (1854)[162], la c.d. algebra booleana, che rende possibile esprimere in forma algebrica le relazioni logiche tra enunciati, utilizzando i valori 0 e 1 per esprimere falsità e verità. L'algebra

[155]Liebniz concepisce la *characteristica universalis* per la prima volta nell'opera giovanile *De arte combinatoria*, per poi sviluppare l'idea nel corso della sua intera produzione filosofica, cfr. in generale G.W. LIEBNIZ, *Scritti filosofici*, a cura di D. O. BIANCA, Torino, 1978.

[156] V. L. COUTURAT, *La logique de Leibniz d'après des documents inédits*, Parigi, 1901, p. 81 ss.

[157] Il filosofo, nello scritto *De arte characteristica ad perficiendas scientias ratione nitentes*(1688) argomentava: «*Quo facto, quando orientur controversiae, non magis disputatione opus erit inter duos philosophos, quam inter duos computistas. Sufficie tenim calamos in manus sur aere sedere que ad abbacos, et sibi mutuo (accito si placet amico) dicere: Calculemus!*», cfr. L. COUTURAT, *op. cit.*, p. 98.

[158]L. COUTURAT, *op. cit.*, 1901, p. 431 ss.

[159]Cfr. *ex multis*, J. NILSSON, *The quest for artificial intelligence cit.*, p. 12.

[160] L'*Ideografia* elaborate da Frege permetteva di esprimere le relazioni tra proposizioni utilizzando lettere e simboli grafici. Il sistema di notazione logica attualmente, elaborato nel XX secolo, in uso è più sintetico e maneggevole. Cfr. F.L.G. FREGE, *Begriffsschrift, eine der arithmetischennachgebildeteFormelsprache des reinenDenkens*, 1879, trad. inglese a cura di S. BAUER-MENGELBERG, *Concept Script, a formal language of pure thought modelled upon that of arithmetic*, in J. VAN HEIJENOORT (A CURA DI), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, Cambridge (US), 1967.

[161]G. BOOLE, *The Mathematical Analysis of Logic*, Cork, 1847.

[162]G. BOOLE, *An Investigation of the Laws of Thought*, Cork, 1854.

booleana sarà fondamentale per lo sviluppo dell'intelligenza artificiale e dell'intera informatica, essendo alla base di ogni linguaggio di programmazione[163]. Ai nostri fini, è interessante notare che né il suo teorizzatore Boole, né l'ideatore dell'*Ideografia* Frege ipotizzarono di applicare le loro innovazioni nel campo della logica per la costruzione di macchine in grado di meccanizzare il ragionamento, come aveva fatto Liebniz più di un secolo prima, e sarebbe nei fatti avvenuto un secolo dopo.

2. **La riflessione filosofica sulla tecnologia e sul rapporto uomo-macchina nell'età contemporanea**

Le prime riflessioni filosofiche riguardanti l'impatto della tecnologia sulla società risalgono al rinascimento, epoca passata alla storia, più di ogni altra, per il fiorire delle abilità creative e innovatrici dell'uomo. Francesco Bacone è considerato il primo autore di rilievo ad affrontare il tema e, nel racconto utopico *La Nuova Atlantide* (1627)[164], offre una visione largamente positiva del fenomeno tecnologico. L'opera narra le vicende dell'isola immaginaria di Bensalem e del popolo che la abita nella concordia e nella prosperità, grazie alle scoperte frutto di un'avanzata cultura scientifica e all'abilità nell'imitare le migliori tecnologie messe a punto dal resto dei paesi civilizzati.

L'entusiasmo tecnologico caratterizzò anche la successiva età dei lumi e rimase l'opinione largamente dominante fino al XIX secolo, quando le conseguenze sociali della prima rivoluzione industriale suscitarono le prime riflessioni critiche[165]. Le terribili condizioni di lavoro nelle fabbriche inglesi dell'epoca, infatti, scalfirono la convinzione che il progresso tecnico portasse sempre a innovazioni positive per la collettività. Nei movimenti operai prevalse, in ogni caso, una visione neutra e strumentale della tecnologia, che, pur avendo perso la caratterizzazione intrinsecamente positiva tipica dell'epoca precedente, si era rivelata portatrice di conseguenze negative solamente in virtù dell'uso, sconsiderato, che aveva scelto di farne chi la governava. Nel *Capitale*, Marx considera gli avanzamenti tecnologici del tempo strettamente connessi ai rapporti di produzione capitalisti – tanto da affermare, eloquentemente, «il mezzo di lavoro schiaccia l'operaio»[166] - ma la sua postura anticapitalista non è anche antitecnologica. La soluzione marxiana alla questione sociale, infatti, consiste nella collettivizzazione dei mezzi di produzione, grazie alla

[163]Cfr. ancora J. NILSSON, *The quest for artificial intelligence cit.*, p. 13-14.

[164] Può indicarsi, tra le molte edizioni, F. BACONE, *La nuova Atlantide*,a cura di P. GUGLIELMONI, Milano, 1997, che raccoglie le versioni in inglese e in latino di Bacone e la traduzione del curatore.

[165]Cfr. M.P.M. FRANSSEN, G.J. LOCKHORST, I. VAN DE POEL, *Philosophy of technology*, in E. N. ZALTA (A CURA DI), *The Stanford Encyclopedia of Philosophy*, 2018, https://plato.stanford.edu/archives/fall2018/entries/technology/ (26 novembre 2021).

[166] K. MARX, *Il Capitale*, I, XIII, IV, a cura di M. L. BOGGERI, Roma, 2016, p. 1109; prima ed. K. MARX, *Das Kapital*, Amburgo,1867.

quale la tecnologia, di per sé neutra, si convertirebbe da strumento di oppressione in strumento di progresso[167]. Lo stesso movimento luddista può dirsi animato da una vera avversione per la tecnologia solo con molti distinguo, posto che la critica violenta dell'introduzione del telaio meccanico che lo caratterizzò aveva l'unico fine di conservare il posto di lavoro dei filatori, ma non è possibile rinvenirvi una compiuta elaborazione filosofica degli effetti del progresso tecnologico sulla società[168]. La prima critica aperta della tecnologia compare, in realtà, in un'opera di narrativa, il romanzo *Erewhon* del britannico Samuel Butler, pubblicato in forma anonima nel 1872[169]. *Erehwon* ripercorre le vicende di una terra in cui le macchine sono bandite ed è un crimine costruirne una, poiché gli abitanti ritengono nocciano alla concordia e allo sviluppo sociale.

Nel '900, la discussione critica sulla tecnologia diventa uno dei temi centrali del dibattito filosofico, in ragione della diffusione sempre più capillare di strumenti tecnologici e dei pericoli per l'intera umanità connessi all'utilizzo bellico del progresso tecnico, culminato nelle bombe di Hiroshima e Nagasaki[170]. Nella prima metà del secolo si distinguono le opere del pensatore tedesco Friedrich Dessauer e dello spagnolo José Ortega y Gasset. Pur da prospettive diverse, questi due filosofi sono i primi a descrivere la tecnica e il progresso come la nuova, totalizzante dimensione in cui si svolge l'esistenza degli esseri umani. Infatti, Dessauer – che, da scienziato, contribuirà al perfezionamento della tecnologia a raggi X – considera, entusiasticamente, la partecipazione al progresso tecnologico una sorta di imperativo categorico kantiano, con cui ogni uomo può riempire la propria vita di significato[171]. Ortega y Gasset, all'opposto, ritiene che le innumerevoli possibilità dischiuse dalla tecnica abbiano privato l'uomo delle categorie della volontà e del desiderio, rendendolo oggetto di un processo non più governabile e aprendo la strada a una moderna forma di nichilismo radicale[172].

Nel secondo dopoguerra, l'idea che la tecnologia sia un fenomeno pervasivo dell'intera realtà e fuori dal controllo dell'uomo è al centro delle riflessioni di Martin Heidegger ne *La Questione della Tecnica* (1953)[173]. Senza prendere posizioni esplicitamente antitecnologiche, Heidegger evidenzia come la tecnologia non possa, in ogni caso, dirsi neutrale, poiché non è più l'essere umano a indirizzarne gli obiettivi, ma è anzi quest'ultima a disegnare l'ambito e la direzione delle azioni

[167] Cfr. in generale J. FALLOT, *Marx e la questione delle macchine*, Firenze, 1971.
[168] V. in generale L. SALVADORI, C. VILLI, *Il luddismo. L'enigma di una rivolta*, Sesto S. Giovanni, 1987.
[169] S. BUTLER, *Erehwon*, a cura di L. DRUDI DEMBY, Milano, 1993; prima ed. ANONIMO, *Erehwon: or over the range*, Londra, 1872.
[170] M.P.M. FRANSSEN, G.J. LOCKHORST, I. VAN DE POEL, *Philosophy of technology cit.*
[171] Cfr. F. DESSAUER, *Filosofia della tecnica* (A CURA DI M. BENDISCIOLI), Brescia, 1945; prima ed. *Philosophie der Technik*, Bonn, 1927. Per un commento v. C. MITCHAM, *Thinking through technology*. *The path between engineering and philosophy*, Chicago, 1994, p. 29-33.
[172] Cfr. in particolare J. ORTEGA Y GASSET, *Meditación de la técnica*, in *Ensimismamiento y alteración. Meditación de la técnica*, Madrid, 1939; per dei commenti v. A. F. GONCALVES JR., *Etica e sociedade tecnológica segundo a filosofia de Ortega y Gasset*, in *Reflexao*, Campinas, 31, 89, p. 25-39; J. L. GONZÁLEZ QUIRÓS, *La meditación de Ortega sobre la técnica y lastecnologíasdigitales*, in *Revista de estudios orteguianos*, 2006, 12-13, p. 95 ss.
[173] M. HEIDEGGER (A CURA DI F. SOLLAZZO), *La questione della tecnica*, Firenze, 2017; prima ed. *Die Fragenach der Technik*, in *Vorträge und Aufsätze*, Pfullingen, 1953.

dell'uomo. Chi, invece, ritiene senza reticenze che le conseguenze negative dello sviluppo tecnologico superino di gran lunga quelle positive è il sociologo francese Jacques Ellul, che ha dedicato al tema gran parte dei suoi studi, tra gli anni '50 e gli anni '80 del '900. Secondo Ellul, la tecnica è la sovrastruttura dominante della società, e le potenzialità dischiuse dalle tecnologie dell'informazione e della comunicazione hanno permesso di unire tutte le sue applicazioni in un unico *sistema tecnico*, che ingloba e condiziona politica ed economia e imprigiona l'uomo nei propri meccanismi. In tale ottica, non vi è spazio per l'umanesimo nella società tecnologica, né è possibile porre rimedio all'incedere di quest'ultima: l'unica possibilità per l'uomo è sviluppare nuove strategie di adattamento e organizzazione, al fine di sopravvivere all'interno di questo scenario[174].

Nel secondo dopoguerra, inoltre, l'indagine sul rapporto tra uomo e tecnica acquisisce una nuova dimensione, a causa dello sviluppo dei primi sistemi di intelligenza artificiale e robotici. Nella letteratura fantascientifica sorge l'ipotesi, variamente declinata, di una "ribellione" delle macchine, destinate a soggiogare l'uomo e divenire la nuova specie intelligente che domina il mondo. L'idea, inizialmente confinata alla narrativa, è entrata nel dibattito filosofico e scientifico con l'avanzamento della tecnologia. Così, le leggi della robotica coniate da Isaac Asimov nei racconti raccolti nel noto volume *Io, robot* del 1950[175] da elemento narrativo si sono convertite nei limiti etici minimi oltre i quali non può spingersi l'azione delle macchine, riconosciuti pressoché universalmente ed entrati in diversi documenti ufficiali[176]. Nel dibattito scientifico, invece, sono state tentate diverse previsioni di quando sarà raggiunta la c.d. *singolarità tecnologica*, ovvero il momento in cui lo sviluppo tecnologico sfuggirà a qualunque controllo dell'essere umano e procederà in autonomia, guidato da un'intelligenza artificiale superiore a quella degli uomini. Il concetto è stato elaborato dal matematico e romanziere statunitense Vernor Vinge in un articolo del 1993 intitolato, appunto, *Technological Singularity*[177], e gli esperti che se ne sono occupati si dividono tra coloro che vi vedono il dischiudersi di inimmaginabili opportunità di benessere e

[174] Cfr. in particolare J. ELLUL, *La technique ou l'enjeu du siècle*, Parigi, 1954; J. ELLUL, *Le Système technician*, Parigi, 1977. V. anche C. MITCHAM, *Thinking through technology cit.*, p. 57-61.

[175] Le prime tre leggi della robotica sono state compiutamente enunciate da Asimov per la prima volta nel racconto *Runaround*, contenuto, appunto, nella raccolta I. ASIMOV, *I, Robot*, New York, 1950. La prima leggerecita: «a robot may not injure a human being or, through inaction, allow a human being to come to harm»; la seconda: «a robot must obey the orders given it by human beings except where such orders would conflict with the First Law»; la terza: «a robot must protect its own existence as long as such protection does not conflict with the First or Second Law». Piùtardi, Asimov stesso vi ha aggiunto una quartalegge, la c.d. "legge zero": «a robot may not harm humanity, or, by inaction, allow humanity to come to harm», enunciata nel romanzo I. ASIMOV, *Robots and Empire*, New York, 1985.

[176] Le leggi di Asimov, ad esempio, sono citate nell'introduzione della Risoluzione del Parlamento Europeo del 16 febbraio 2017 recante *raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica* (2015/2103(INL)). Per alcune considerazioni sull'influenza degli scritti di Asimov cfr. R.CLARKE,*Asimov's laws of robotics: implications for information technology*, in *Computer*, Dec. 1993, p. 53-61 (pt. I) e Jan. 1994, p. 57-66 (pt. II).

[177] V. VINGE, *The Coming Technological Singularity: How to Survive in the Post-Human Era*, in *Vision-21 Interdisciplinary Science and Engineering in the Era of Cyberspace - Proceedings of a symposium cosponsored by the NASA Lewis Research Center and the Ohio Aerospace Institute and held in Westlake (Ohio)*, 1993, p. 11-22.

progresso e coloro che, all'opposto, vi scorgono un pericolo per la stessa sopravvivenza del genere umano[178]. A prescindere dalle diverse interpretazioni e previsioni, deve evidenziarsi che, allo stato dell'arte, lo sviluppo di un'intelligenza artificiale generale in grado di prendere il sopravvento pare ben distante dal realizzarsi[179]. Al contempo, le applicazioni dell'intelligenza artificiale che già esistono, nonostante non siano così avanzate da far apparire prossimo il raggiungimento della singolarità, pongono una varietà di interrogativi etici ben più concreti e urgenti. La riflessione contemporanea sull'intelligenza artificiale preferisce concentrarsi sulle possibili soluzioni a tali problemi e ha in buona parte abbandonato le riflessioni d'ampio respiro sul rapporto uomo-macchina.

Si tratta, peraltro, di una tendenza che ha investito l'intera filosofia della tecnologia degli ultimi decenni. Infatti, la pervasività del fenomeno tecnologico e l'impossibilità, per l'uomo, di concepirsi al di fuori di esso – come visto, al centro della riflessione nel XX secolo – sono date per acquisite da molti pensatori contemporanei, che preferiscono concentrarsi sulle possibili strategie di adattamento all'ambiente tecnologico e cercare soluzioni ai problemi etici che si generano al suo interno. È il caso, ad esempio, di correnti come la statunitense *analytical philosophy of technology*, che indaga gli elementi necessari, anche dal punto di vista del *design*, per un'implementazione della tecnologia rispettosa delle prerogative dell'essere umano[180], o la filosofia dell'informazione del XXI secolo, che assume come oggetto di studi la vita nell'*Infosfera*, il nuovo ambiente digitale composto da reti di macchine interconnesse[181].

3. **(Alcuni) dilemmi etici dell'intelligenza artificiale**

Giunti a questo punto, è utile dare conto, senza animo di completezza, di alcune delle questioni etiche più spinose sollevate dalle correnti applicazioni dell'intelligenza artificiale. Infatti, i capitoli

[178]Metteva in guardia dai pericoli connessi allo sviluppo tecnologico l'ingegnere informatico statunitense Bill Joy, che in un celebre articolo del 2000 affermava: «Our most powerful 21st-century technologies — robotics, genetic engineering, and nanotech — are threatening to make humans an endangered species», cfr. B. JOY, *Why the future doesn't need us*, in *Wired*, 2000, https://www.wired.com/2000/04/joy-2/ (25 novembre 2021). Lo stesso V. VINGE, *The coming technological singularity cit.*, R. KURZWEIL, *The age of intelligent machines*, Cambridge, 1990; *The age of spiritual machines*, Cambridge, 1998; *The singularity is near*, New York, 2005 e N. BOSTROM, *Superintelligence: paths, dangers, strategies*, Oxford, 2014 vedono invece nella singolarità tecnologica l'inizio di una nuova era di potenziale prosperità per l'essere umano, qualora ne fossero scongiurati i rischi. Tra gli studi sulla singolarità tecnologica deve citarsi anche il recente e molto noto J. LOVELOCK, *Novacene: the coming age of hyperintelligence*, Londra, 2019, che vede nella diffusione di machine intelligenti l'inizio di una nuova era, portatrice per l'uomo, sia di pericoli che di potenziale prosperità, nella storia del pianeta. Si deve specificare che, in ogni caso, la stessa ipotesi della singolarità è avversata da molti studiosi, che la ritengono irrealizzabile: cfr. ad es. *The reality club: one half of a Manifesto*, in *Edge.org*, 11 ottobre 2000.

[179] V. ancora S. BRINGSJORD, N.S. GOVINDARAJULU, *Artificial Intelligence cit.*; R. FJELLAND, *Why general artificial intelligence will not be realized*, in *Humanities and social sciences communications*, 7, 10, 2020.

[180] Cfr. *inter alia* M.P.M. FRANSSEN, G.J. LOCKHORST, I. VAN DE POEL, *Philosophy of technology cit.*

[181] Si rimanda in primo luogoall'opera di Luciano Floridi, in particolare, tra gli altri, L. FLORIDI, *The philosophy of information*, Oxford, 2011; *The ethics of information*, Oxford, 2013; *The fourth revolution – How the Infosphere is reshaping human reality*, Oxford, 2014; *The logic of information*, Oxford, 2019.

seguenti saranno dedicati al precipitato giuridico di tali questioni e agli strumenti con cui il diritto potrà farvi fronte, al fine di minimizzare i rischi rappresentati dalle tecnologie intelligenti senza, per questo, rinunciare ai loro benefici. È bene, quindi, offrire sinteticamente una panoramica dei nodi problematici su cui si concentra la discussione tra gli esperti del settore:

- **L'impatto sul mondo del lavoro**[182]**.** Una delle preoccupazioni più comuni riguarda l'ipotesi che la diffusione dell'intelligenza artificiale nei processi produttivi porti a una radicale diminuzione dei posti di lavoro complessivi, a causa della sostituzione di molti operatori umani con sistemi automatizzati. In realtà, si tratta di timori che hanno accompagnato diverse innovazioni tecnologiche del passato, rivelandosi spesso infondati: l'avvento di nuovi macchinari ha portato a una trasformazione del lavoro e non a una sua diminuzione, e la scomparsa di alcune figure professionali è stata accompagnata dalla nascita di altre, maggiormente qualificate. Ciò, comunque, non rende di per sé meno urgente il problema, posta la velocità con cui le tecnologie intelligenti si diffondono nei meccanismi produttivi, sostituendo manodopera che, spesso, è priva di una formazione adeguata per svolgere le nuove mansioni che si rendono necessarie. Adeguare il sistema di formazione e ricollocamento dei lavoratori al mutato scenario economico e tecnologico sarà, probabilmente, una delle sfide maggiori poste dall'avvento dell'intelligenza artificiale nei paesi avanzati.
- **Privacy, profilazione e sorveglianza**[183]. Com'è noto, i sistemi di *machine* e *deep learning* si basano sull'elaborazione di moli di dati sempre più grandi. Quando il loro impiego coinvolge l'utilizzo di dati personali si pongono problemi specifici dal punto di vista della riservatezza e del controllo su questi ultimi, perché l'analisi di dati aggregati permette profilazioni particolarmente accurate delle caratteristiche e preferenze individuali. Si tratta di informazioni utilizzabili, ad esempio, per finalità commerciali o per rendere più efficace la propaganda politica, talvolta con disturbanti conseguenze. Inoltre, pone pesanti interrogativi l'utilizzo di tecnologie intelligenti con finalità di sorveglianza, pubblica

[182] Sul tema in generale cfr., da vari punti di vista, J. RIFKIN, *La fine del lavoro. Il declino della forza lavoro globale e l'avvento dell'era post-mercato*, Milano, 1995; M. FORD, *Rise of the Robots: Technology and the Threat of a Jobless Future*, New York, 2015; B. ARTHUR, *The second economy*, in *McKinsey Quarterly*, Oct. 2011, https://mck.co/3BodpaQ (3 dicembre 2021); D. H. AUTOR, *Why are there still so many jobs? The history and future of workplace automation*, in *Journal of Economic Perspectives*, 29, 2015, p. 3; M. CASELLI ET AL., *Stop worrying and love the robot: an activity-based approach to assess the impact of robotization on employment dynamics*, in *GLO Discussion Paper*, Essen, 2021, 802, p. 30 ss.; M. BORZAGA, *Le ripercussioni del progresso tecnologico e dell'Intelligenza Artificiale sui rapporti di lavoro in Italia*, in *DPCE online*, 1, 2022, p. 393-403.

[183] Cfr. ad es. A. MANTELERO, *Artificial Intelligence and Data Protection: Challenges and Possible Remedies – Report for the Council of Europe consultative committee of the Convention 108* (T-PD(2018)09Rev), 2018; G. FINOCCHIARO, *Intelligenza artificiale e protezione dei dati personali*, in *Giurisprudenza italiana*, 2019, p. 1670-1678. Sui rischi connessi a profilazione e sorveglianza v. anche S. ZUBOFF, *The age of surveillance capitalism: the fight for a human future at the new frontier of power*, Londra, 2019; L. EFREN RIOS VEGA, L. SCAFFARDI, I. SPIGNO (a cura di*), I diritti fondamentali nell'era della Digital Mass Surveillance*, Napoli, 2021.

sicurezza e repressione del crimine. Infatti, l'uso simultaneo di algoritmi predittivi del comportamento e strumenti di sorveglianza di ultima generazione, come imoderni sistemi di riconoscimento facciale,sembra erodere sempre di più la sfera privata del cittadino.

• **Opacità e controllo umano sull'intelligenza artificiale**[184]. Alcune tecnologie intelligenti, in particolare quelle basate sul*deep learning* e altre applicazioni avanzate dell'apprendimento automatico, rendono estremamente difficile per l'operatore umano ricostruire l'iter logico con cui giungono ai loro risultati. Questo può rivelarsi particolarmente problematico quando l'intelligenza artificiale sia impiegata in settori particolarmente delicati, nei quali l'operatore umano si troverebbe di fronte a un'indicazione algoritmica della quale non potrebbe decifrare appieno le ragioni. Inoltre, l'impiego di sistemi dotati di autonomia in contesti che presentano un'ineliminabile area di rischio per gli esseri umani, come le auto a guida autonoma o la robotica industriale, pone l'interrogativo morale dell'opportunità dell'utilizzo di strumenti di fatto sottratti al controllo dell'essere umano in situazioni intrinsecamente pericolose, oltre a quello etico e giuridico del collocamento della responsabilità qualora tali rischi si realizzino.

• **Data bias e algorithmic bias**[185]. I sistemi di *machine learning* e *deep learning* che formulano previsioni, decisioni e valutazioni presentano il rischio di output errati a causa di difetti nella formazione dei dataset su cui sono stati allenati, in cui alcune variabili o interi gruppi potrebbero essere sotto o sovrarappresentati. Il carattere intrinsecamente statistico del funzionamento di tali algoritmi, inoltre, rende ineliminabile la presenza di un margine d'errore nei loro risultati, una circostanza che può rivelarsi di difficile gestione a causa dell'opacità che spesso li caratterizza.

• **Utilizzo bellico dell'intelligenza artificiale**[186]. Negli ultimi due decenni sono state sviluppate diverse tecnologie basate sull'intelligenza artificiale destinate all'uso bellico. Le

[184]In via generale e da distinti punti di vista, cfr. F. PASQUALE, *The Black-Box Society: The Secret Algorithms That Control Money and Information*, Harvard, 2016; R. V. YAMPOLSKIY, *Unexplainability and incomprensibility of artificial intelligence*, 2019, arXiv:1907.03869; G. VILONE, E. LONGO, *Explainable artificial intelligence: a systematic review*, 2020, arXiv:2006.00093; A. PAEZ, *The Pragmatic Turn in Explainable Artificial Intelligence (XAI)*, in *Minds and Machines*, 29, 2019, p. 441-459; M. L. JONES, *The right to a human in the loop: Political constructions of computer automation and personhood*, in *Social Studies of Science*, 47, 2, 2017, p. 216-239; MONREALE A., *Rischi etico-legali dell'intelligenza artificiale*, in *DPCE Online*, 3, 2020, p. 3391-3398.

[185]Cfr. B. FRIEDMAN, H. NISSENBAUM, *Bias in computer systems*, in *ACM Transactions on Information Systems*, 14, 3, 1996, https://doi.org/10.1145/230538.230561; R. DOBBE, S. DEAN, T. GILBERT, N. KOHLI, *A Broader View on Bias in Automated Decision-Making: Reflecting on Epistemology and Dynamics*, 2018, arXiv:1807.00553; oltreagliesempiraccolti in C. O'NEILL, *Weapons of math destruction: how big data increases inequality and threatens democracy*, Largo (USA), 2016. Tra gli esempi più noti di *bias* algoritmo, può citarsi fin d'ora il *software* COMPAS, usato a supporto di valutazioni di pericolosità di imputati e condannati nel sistema penale statunitense ed esposto a un *bias* a danno della comunità afroamericana. Cfr. *infra*, p. 153 ss.

[186] In generale, D. AMOROSO, *Jus in bello and jus ad bellum arguments against autonomy in weapons systems: a re-appraisal*, in *Questions on International Law,* Zoom in 43, 2017, p. 5-31; M. SASSÓLI, *Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified*, in

preoccupazioni sollevate dai c.d. LAWS (acronimo di *Lethal Autonomous Weapons Systems*) riguardano, in primo luogo, il controllo umano sul comportamento di tali sistemi, la possibilità di errori fatali nelle loro valutazioni e la paura che il loro impiego porti a una svalutazione della vita umana, disumanizzando ulteriormente il contesto bellico.

International law studies, 90, 2014, p. 308-340; P. ASARO, *On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making*, in *International Review of the Red Cross*, 2012, 94, p. 687-709; F. CHESINI, *"Terminator scenario"? Intelligenza artificiale nel conflitto armato: "lethal autonomous weapons systems" e le risposte del diritto internazionale umanitario*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2020, p. 441-471.

Regolare l'intelligenza artificiale: lo stato dell'arte del "diritto dell'IA"

1. Un diritto per l'intelligenza artificiale: il rischio di una nuova *law of the horse*?

Sono passati più di vent'anni da quando il giudice Frank Easterbrook, in una conferenza all'università di Chicago dedicata alla "Law of Cyberspace", gelò la platea, composta da esperti di tale nuova, emergente disciplina, affermando che una "Law of Cyberspace" non aveva più dignità scientifica di un'ipotetica "law of the horse"[187]. L'argomento non può non suscitare divertita ammirazione per semplicità e, allo stesso tempo, brillantezza: Easterbrook spiegò che vi sono norme che riguardano il commercio dei cavalli, la loro macellazione, le scommesse sulle loro corse o le cure veterinarie di cui hanno bisogno, ma non per questo, nelle università, si insegna un corso di diritto del cavallo. L'intervento voleva essere un monito a non scivolare in pericolose forme di dilettantismo interdisciplinare, e a non sottovalutare le possibilità di adattamento delle categorie giuridiche tradizionali anche ai contesti più innovativi. È ugualmente famosa la replica di Lawrence Lessig nel volume *Code and other laws of the cyberspace*[188], risalente a tre anni dopo e a cui la storia recente sembra aver dato ragione. La ragione essenziale per cui Lessig dissentiva da Easterbrook era che il nuovo spazio digitale, per l'importanza e la capillarità che stava acquisendo, sembrava destinato ad avere un impatto sul diritto come fenomeno globale e non solo su alcune, particolari discipline, a differenza dei cavalli. Lessig, in realtà, si spingeva a dire che il diritto, e i valori posti a suo fondamento, erano posti in pericolo dall'avvento del cyberspazio, nel quale sarebbero potuti sorgere strumenti di regolazione e poteri alternativi a quelli statali, grazie alla rete. Secondo la sua tesi, questo insieme di circostanze rendeva necessario uno studio sistematico e unitario del fenomeno digitale dal punto di vista del diritto.

Due decenni dopo, l'intelligenza artificiale suscita un dibattito molto simile. Diverse voci, infatti, considerano necessaria una regolazione unitaria del fenomeno, un "diritto dell'intelligenza artificiale"[189]. Parallelamente, altri mettono in guardia dal rischio di dar vita a una nuova *law of the horse*[190]. In realtà, all'obiezione si può rispondere facilmente con gli stessi argomenti usati da Lessig vent'anni fa: la pluralità di applicazioni già esistenti dell'intelligenza artificiale – per tacere

[187] F. H. EASTERBROOK, *Cyberspace and the law of the horse*, in *1996 University of Chicago Legal Forum*, 1996, p. 207-216.

[188] L. LESSIG, *Code and other laws of cyberspace*, New York, 1999.

[189] *Ex multi s*cfr., da vari punti di vista, M. C. BUITEN, *Towards intelligent regulation of artificial intelligence*, in *European Journal of Risk Regulation*, 10, 1, 2019, p. 41-59; E. STRADELLA, *La regolazione della robotica e dell'intelligenza artificiale: il dibattito, le proposte, le prospettive. Alcuni spunti di riflessione*, in *Media Laws*, 1, 2019, p. 73-92; A. D'ALOIA (A CURA DI), *Intelligenza artificiale e diritto: come regolare un mondo nuovo*, Milano, 2020; A.A.V.V., *Il diritto comparato dell'intelligenza artificiale*, in *Diritto Pubblico Comparato ed Europeo*, numero monografico 1, 2022.

[190] Si vedano, ad esempio, le considerazioni di C. DOCTOROW, *Why it is not possible to regulate robots*, in *The Guardian*, 2 aprile 2014; cfr. anche N. M. RICHARDS, W. D. SMART, *How should the law think about robots?*, in R. CALO, A. M. FROOMKIN, I. KERR (A CURA DI), *Robot Law*, Cheltenham, 2016, p. 3-22.

di quelle, solo in parte immaginabili, che riservano gli anni a venire – sta imponendo trasformazioni pressoché in ogni ambito della vita economico-sociale delle nostre collettività, oltre che in numerosi aspetti della sfera personale e relazionale del singolo individuo. Mentre ciò avviene, si presentano questioni sempre più urgenti, come quelle elencate al paragrafo precedente. Spetta, ovviamente, in primo luogo al diritto trovarvi risposta, adattando i propri strumenti. L'intelligenza artificiale, quindi, ha un impatto sul fenomeno giuridico nella sua globalità, a cominciare dai valori che ne costituiscono il fondamento e dai diritti che caratterizzano la nostra tradizione costituzionale, come si cercherà di analizzare nel seguito di questo lavoro. Parlare di un nuovo "diritto dell'intelligenza artificiale" sembra, perciò, pienamente legittimo.

Più in generale, deve rilevarsi che la questione ha, in fondo, scarso rilievo pratico. Definire se si possa parlare con compiutezza di un "diritto dell'intelligenza artificiale" o se si tratti, invece, di interpretare un nuovo fenomeno tecnologico con le categorie giuridiche tradizionali, può essere un affascinante interrogativo teorico, analogamente a quanto avvenuto col cyberspazio. Dal punto di vista sostanziale, però, rimane in ambo i casi la necessità di trovare una risposta da parte del diritto alle numerose questioni aperte dall'avvento dell'intelligenza artificiale, che rappresentano alcune tra le principali sfide giuridiche del presente e del futuro. Se trovare gli adeguati strumenti per farvi fronte porti o meno alla nascita di una nuova disciplina giuridica è un tema, in ultima analisi, di rilevanza limitata.

2. La regolazione dell'intelligenza artificiale nei paesi industrializzati e a livello sovranazionale: piani strategici, documenti di *soft-law* e prime ipotesi di strumenti di *hard-law*

Negli ultimi anni, le principali potenze economiche mondiali hanno prodotto piani strategici per orientare lo sviluppo dell'intelligenza artificiale nei prossimi decenni. L'obiettivo di queste programmazioni è far coincidere il più possibile il progresso tecnologico con la visione di interesse nazionale (o sovranazionale, nel caso dell'Unione Europea) di volta in volta adottata. Inoltre, tutti i piani riservano grande attenzione allo scenario internazionale, sul quale ciascuna potenza ambisce a guadagnare posizioni nella corsa allo sviluppo tecnologico, o, nel caso degli Stati Uniti, mantenere il proprio primato. È di particolare interesse che, pur con differenti declinazioni, preoccupazioni di carattere etico e sociale siano esplicitate in ognuno di questi progetti, anche se provenienti da paesi non democratici, come la Repubblica Popolare Cinese.

Per quanto riguarda gli Stati Uniti, una prima *American AI Initiative*è stata emanata nel febbraio 2019, con *Executive Order* presidenziale di Donald Trump[191]. Il piano è stato riorganizzato, con il

[191] Exec. Order n. 13859 of Feb 11, 2019, *Maintaining American leadership in artificial intelligence*.

passaggio all'amministrazione Biden, dal *National Artificial Intelligence Initiative Act*[192], in vigore dal gennaio 2021 e approvato con il voto trasversale di democratici e repubblicani. I principi ispiratori della strategia statunitense sono, comunque, rimasti pressoché gli stessi: sia l'ordine presidenziale del 2019 che l'atto legislativo del 2021 dichiarano l'obiettivo di conservare il proprio vantaggio e favorire uno sviluppo tecnologico accompagnato da una proficua formazione continua dei lavoratori e dal mantenimento dei più alti standard di qualità e sicurezza della produzione[193]. Inoltre, entrambi i piani prendono in considerazione l'impatto dell'IA sulla società e sul diritto: l'*Artificial Intelligence Initiative Act* sancisce l'intenzione degli Stati Uniti di *«lead the world in the development and use of trustworthy artificial intelligence systems in the public and the private sector»*[194], mentre la precedente *AI Initiative* del 2019 poneva l'accento sulla necessità di proteggere i valori fondanti degli Stati Uniti, come privacy, libertà e diritti civili[195]. Peraltro, l'entrata in vigore del *National Artificial Intelligence Initiative Act* è stata accompagnata dalla creazione di diversi enti amministrativi al fine di assicurarne la corretta implementazione, al vertice dei quali è stato posto il *National Artificial Intelligence Initiative Office*[196], con compiti di coordinamento a livello federale.

L'Unione Europea, invece, ha reso pubblica la sua strategia per l'intelligenza artificiale nel 2018, con la Comunicazione della Commissione al Parlamento Europeo e al Consiglio *L'intelligenza artificiale per l'Europa*[197]. Il piano si basa su tre pilastri: consolidare l'Unione Europea come potenza tecnologica e incoraggiare l'uso dell'intelligenza artificiale nel settore pubblico e privato; governare le trasformazioni socioeconomiche che essa è destinata a causare; costruire un sistema di garanzie etiche e legali per lo sviluppo di un'intelligenza artificiale affidabile e *human-centric*[198]. Nello stesso anno, per l'implementazione della strategia è stato emanato il *Coordinated Plan on Artificial Intelligence*, frutto di una lunga consultazione tra la Commissione e gli Stati membri, sottoscritto anche da Norvegia e Svizzera e sottoposto a revisione annuale[199]. Con il documento, l'Unione e gli Stati membri hanno definito una serie di azioni comuni, al fine di accrescere gli investimenti nel settore, favorire la disponibilità di dati con cui sviluppare tecnologie intelligenti e aumentare il numero di persone con le necessarie competenze tecnologiche. Inoltre, sono stati

[192] National Artificial Intelligence Initiative Act, H.R. 6216, 116th Cong. (2020).
[193] Cfr. in particolare le sec. 1-2 dell'Executive Order del 2019 e la sec. 101 Titolo I del National Artificial Intelligence Initiative Act del 2020.
[194] L'affermazione è contenuta alla sec. 101 lett. a) c. 2 del Titolo I del testo.
[195] La sec. 1 lett. d) dell'Executive Order recita: «The United States must foster public trust and confidence in AI technologies and protect civil liberties, privacy, and American values in their application in order to fully realize the potential of AI technologies for the American people».
[196] Cfr. il sito ufficiale messo a punto dal governo degli Stati Uniti: https://www.ai.gov/naiio/ (7 novembre 2021).
[197] Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato Economico e Sociale Europeo e al Comitato delle Regioni, *L'intelligenza artificiale per l'Europa*, 24 aprile 2018, COM(2018) 237 final.
[198] Cfr. in particolare il par. 3 della Comunicazione: *«La strada da seguire: un'iniziativa dell'Ue per l'IA».*
[199] Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato Economico e Sociale Europeo e al Comitato delle Regioni, *Coordinated plan on artificial intelligence*, 7 dicembre 2018, COM(2018) 795 final (Annex I).

individuati alcuni settori nei quali l'investimento nell'intelligenza artificiale è considerato prioritario: salute, mobilità, sicurezza, energia, produzione industriale e servizi finanziari[200].

Nel caso europeo, poi, l'interesse per la dimensione etico-sociale delle tecnologie intelligenti e per i potenziali rischi connessi alla loro diffusione è particolarmente evidente. Infatti, l'approccio della Commissione è rivolto a rendere il mercato europeo il terreno d'elezione per lo sviluppo di un'intelligenza artificiale genuinamente affidabile e *human centric*. Ciò si evince, in primo luogo, dalle esplicite dichiarazioni d'intenti in tal senso. Il *Coordinated Plan on Artificial Intelligence* del 2018, ad esempio, affermava: «*the ambition is for Europe to become the world-leading region for developing and deploying cutting-edge, ethical and secure AI, promoting a human-centric approach in the global context*»[201]. L'indicazione è stata ribadita nella revisione del 2021, che enuncia l'obiettivo di una «*global leadership on trustworthy AI*»[202]. In secondo luogo, l'attenzione della Commissione verso lo sviluppo di un'intelligenza artificiale affidabile è resa palese dal lavoro del Gruppo di Esperti di Alto Livello da essa nominato come organo consultivo per la strategia europea sull'intelligenza artificiale. I lavori del gruppo, infatti, hanno riguardato prevalentemente le cautele necessarie per ottenere questo risultato: oltre alle già più volte menzionate *Ethics Guidelines for Trustworthy AI*, sono di particolare rilievo, tra i documenti elaborati, le *Policy and Investment Recommendations for Trustworthy AI*[203]e la *Assessment List for Trustworthy AI*[204], che mirano, attraverso la definizione di indicazioni pratiche e di un procedimento di verifica delle caratteristiche di un sistema intelligente, a dare concretezza alle linee guida.

Sempre sul continente europeo, un ruolo autonomo è stato assunto dal Regno Unito, che, al momento dell'elaborazione della strategia europea appena descritta, aveva già avviato la procedura di uscita dall'Unione. Il governo del Paese, così, nel 2018 ha lanciato il proprio piano strategico per l'intelligenza artificiale, con la pubblicazione dell'*AI Sector Deal*, un *policy paper* con cui sono stati presentati i principali obiettivi britannici nel breve e medio termine[205]. Il documento dichiara l'intenzione di consolidare il Regno Unito tra le potenze mondiali nel campo dell'intelligenza

[200] Cfr. *Coordinated plan on artificial intelligence cit.*, p. 2: «The coordinated plan will maximise the benefits of AI for all Europeans by fostering the development of trusted AI that corresponds to European ethical values, and citizens' aspirations. Europe will progressively increase its effort in public interest areas such as healthcare, transport, security, education and energy as well as in other areas such as manufacturing and financial services (including through blockchain)».

[201] Si veda la prima pagina del piano.

[202] Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato Economico e Sociale Europeo e al Comitato delle Regioni, 21 aprile 2021, COM(2021) 205 final (Annex – *Coordinated plan on artificial intelligence 2021 review*) p. 2: «The 2021 review of the Coordinated Plan is the next step – it puts forward a concrete set of joint actions for the European Commission and Member States on how to create EU global leadership on trustworthy AI».

[203]HIGH LEVEL EXPERT GROUP ON AI della Commissione UE, *Policy and Investment Recommendations for Trustworthy AI*, 26 giugno 2019.

[204]HIGH LEVEL EXPERT GROUP ON AI della Commissione UE, *Policy and Investment Recommendations for Trustworthy AI*, 17 luglio 2020.

[205] UK GOV. – DEPT. FOR DIGITAL, CULTURE, MEDIA AND SPORT, *AI sector deal*, 26 aprile 2018, https://bit.ly/3DDOBwG (7 dicembre 2021).

artificiale, con un'attenzione specifica per l'attrazione di talento e risorse dall'esterno. Il piano, infatti, prevede forti investimenti nella formazione di alto livello, particolari incentivi per il trasferimento nel Paese di investimenti stranieri, e la creazione di contesti economici e fiscali particolarmente favorevoli per la creazione di imprese operanti nel settore, al fine di «*be the world'smost innovative economy*»[206]. Parallelamente, sono stati creati l'*AI Council*[207], un gruppo di esperti provenienti soprattutto dall'industria tecnologica e dal mondo accademico, incaricato di supportare la messa in atto dell'*AI Sector Deal*, il *Center for Data Ethics and Innovation*[208], allo scopo di individuare le *best practice* per lo sviluppo di un'intelligenza artificiale affidabile e sicura, e un apposito *Office for Artificial Intelligence* nei ministeri di industria e sviluppo economico e di digitale, cultura, media e sport[209]. Nel settembre 2021, inoltre, il governo britannico ha presentato la *National AI strategy*[210], un dettagliato piano industriale elaborato sulla base del menzionato *Sector Deal* del 2018, con l'obiettivo di orientare lo sviluppo del settore dell'intelligenza artificiale per i dieci anni successivi. Il documento, riprendendo quanto già previsto nel piano di tre anni prima, dichiara l'obiettivo di «make Britain a global AI superpower»[211] e individua tre aree di lavoro fondamentali: progettare investimenti a lungo termine nel campo dell'intelligenza artificiale; limitare la diseguaglianza, garantendo che i benefici della rivoluzione tecnologica coinvolgano ogni settore economico e area del paese; governare in modo efficace l'intelligenza artificiale, favorendo lo sviluppo tecnologico e proteggendo, al contempo, l'interesse pubblico[212].

Per quanto riguarda il continente asiatico, è di particolare interesse il comportamento della Cina. Nel 2017 il gigante asiatico ha lanciato il suo *New Generation Artificial Intelligence Development Plan*[213], con il fine dichiarato di fare della Cina il leader mondiale nel settore dell'intelligenza artificiale nel 2030, per la cui implementazione è stato nominato un apposito gruppo di esperti con funzione di consiglieri del governo centrale, l'*AI Strategy Advisory Committee*. Il piano stabilisce l'obiettivo che il Paese diventi, entro quella data, il centro mondiale dell'innovazione tecnologica

[206] Cfr. in particolare il par. 6 del Piano: *«Ideas»*.
[207] Per composizione e funzionamento cfr. https://www.gov.uk/government/groups/ai-council (7 novembre 2021).
[208] Cfr. https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation (7 novembre 2021).
[209] Cfr. https://www.gov.uk/government/organisations/office-for-artificial-intelligence (7 novembre 2021).
[210] UK GOVERNMENT, *National artificial intelligence (AI) strategy*, 22 settembre 2021, https://bit.ly/3dxZCF9 (7 novembre 2021).
[211] Si vedano le p. 4-5 del piano strategico, intitolate per l'appunto «Ourten-year plan to make Britain a global AI superpower».
[212] I progressi nell'attuazione del piano, un anno dopo, sono stati illustrati nel documento UK GOVERNMENT, *National AI Strategy – AI Action Plan*, diffuso il 18 luglio 2022, https://bit.ly/3CWpyXfc (23 agosto 2022).
[213] CONSIGLIO DI STATO DELLA REPUBBLICA POPOLARE CINESE, doc. n. 35, 8 luglio 2017. Una traduzione inglese a cura della *Foundation for Law and International Affairs* è disponibile al link: https://flia.org/notice-state-council-issuing-new-generation-artificial-intelligence-development-plan/ (7 novembre 2021). Per un commento v. F. WU, C. LU, M. ZHU, H. CHEN, J. ZHU, L. LI, M. LI, Q. CHEN, X. LI, X. CAO, Z. WANG, Z. ZHA, Y. ZHUANG, Y. PAN, *Towards a new generation of artificial intelligence in China*, in *Nature Machine Intelligence*, 2, 2020, p. 312-316. Cfr. inoltre I. CARDILLO, *Disciplina dell'intelligenza artificiale e intelligentizzazione della giustizia*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2022, p. 139-167.

nel campo, e fissa in un trilione di yuan (circa 128 miliardi di Euro) il valore complessivo che l'industria dell'intelligenza artificiale dovrà raggiungere[214]. Parallelamente, è previsto lo sviluppo e la codificazione di un sistema di norme giuridiche e di standard etici che regolino l'intelligenza artificiale, che per il 2030 dovrà essere ampiamente consolidato, al fine di poterlo proficuamente aggiornare in funzione delle sfide che dovessero via via presentarsi[215]. L'avvio del piano strategico è stato accompagnato dalla nomina, da parte del governo centrale, di alcune aziende quali eccellenze nazionali, a cui fornire supporto specifico al fine di consolidare la loro posizione in determinati segmenti di mercato, fra le quali spiccano le note Alibaba, per il settore delle *smart cities*, Baidu, per le auto a guida autonoma, e Tencent, per la diagnostica per immagini[216].

Anche le altre due potenze digitali asiatiche, Giappone e Corea del Sud, hanno avviato piani di sviluppo a livello nazionale relativi all'intelligenza artificiale, al fine di non perdere il ruolo di primo piano che attualmente rivestono in alcuni settori del mercato tecnologico. Per quanto riguarda il Giappone, i progetti del Paese sull'intelligenza artificiale sono strettamente connessi ai problemi percepiti come più urgenti nella società: la scarsa crescita economica e il crescente invecchiamento della popolazione. Il governo giapponese ha coniato il concetto di *società 5.0*, dandosi l'obiettivo di creare *«a human-centred society that balances economic advancement with the solution of social problems by a system that highly integrates cyberspace and physical space»*[217]. Nel quadro di questo generale programma di sviluppo è stato nominatolo *Strategic Council for AI Technology*, un organo interministeriale tra i dicasteri degli Interni, dell'Economia e della Comunicazione, che nel 2017 ha presentato l'*Artificial Intelligence Technology Strategy*[218]. Il piano identifica tre aree prioritarie in cui concentrare investimenti e sforzi industriali: sanità (in connessione al menzionato problema dell'invecchiamento), mobilità e produzione industriale (il Paese mira a conservare l'attuale primato mondiale nell'esportazione di robot industriali). Particolare attenzione, inoltre, è posta nell'attrarre talento dall'estero. Al documento del 2017 ha fatto seguito, nel 2019, la pubblicazione di un'*AI Strategy* aggiornata, che pone l'attenzione, in particolare, sullo sviluppo di infrastrutture adeguate, anche dal punto di vista della sicurezza informatica, all'accumulo di moli di

[214] Nella citatatraduzione inglese a curadella FLIA, p. 6: «make core industry scale of artificial intelligence be more than 1 Trillion yuan, driving the scale of related industries be more than 10 trillion yuan».
[215] «Strengthen the research on legal, ethical and social issues related to AI, and establish laws, regulations and ethical frameworks to ensure the healthy development of AI» *Ibidem*, p. 25.
[216] Cfr. H. ROBERTS, J. COWLS, J. MORLEY, M. TADDEO, V. WANG, L. FLORIDI, *The Chinese approach to artificial intelligence: an analysis of policy, ethics and regulation*, in *AI & Society*, 2021, p. 61.
[217] La citazione proviene dalla pagina relativa alla *società 5.0* messa a punto, in lingua inglese, dallo stesso governo giapponese: https://www8.cao.go.jp/cstp/english/society5_0/index.html (7 novembre 2021).
[218] Cfr. *Japan pushing ahead with Society 5.0 to overcome chronic social challenges*, *UNESCO Science Report*, 21 febbraio 2019, https://en.unesco.org/news/japan-pushing-ahead-society-50-overcome-chronic-social-challenges (7 novembre 2021); una traduzione inglese non ufficiale dell'*Artificial Intelligence Technology Strategy*è disponibile a questo link: https://bit.ly/3DApieN (7 novembre 2021).

dati sempre crescenti e aggiunge ai menzionati settori chiave i campi dell'agricoltura, dello sviluppo delle zone svantaggiate del Paese e delle tecnologie per il contrasto dei disastri naturali[219].

Anche la Corea del Sud ha presentato il suo piano strategico per l'intelligenza artificiale nel 2017, con la pubblicazione, a cura del Ministero dello Sviluppo Scientifico e Tecnologico, del *Mid-to-Long-Term Master Plan in Preparation for the Intelligent Information Society: Managing the Fourth Industrial Revolution*[220]. Il progetto si inquadra in una strategia di sviluppo più ampia, coordinata dal Presidential Committee on the Fourth Industrial Revolution, appositamente nominato dal governo al fine di implementare il progetto di una *I-Korea 4.0*, una società basata su uno sviluppo tecnologico orientato alla crescita economica e alla soluzione dei problemi della società[221]. Al fine di raggiungere tale obiettivo, per quanto riguarda le applicazioni dell'intelligenza artificiale il comitato ha individuato 12 settori strategici verso cui indirizzare gli sforzi economici e produttivi: sanità, settore manifatturiero, veicoli a guida autonoma, energia, finanza, pesca e agricoltura, difesa, sicurezza, ambiente, welfare, trasporti e *smart cities*[222].

Accanto ai piani di sviluppo strategico e industriale messi a punto dai paesi avanzati, proliferano le dichiarazioni che elencano i principi per orientare lo sviluppo tecnologico verso un'intelligenza artificiale etica, affidabile e realmente antropocentrica. Si tratta di documenti elaborati, nella maggior parte dei casi, da organismi internazionali, gruppi di esperti formati ad hoc, centri di ricerca e organizzazioni non governative, nessuno dei quali, ad oggi, è legalmente vincolante.

Limitandosi a menzionare, senza animo di completezza, i contributi più rilevanti – l'ammontare totale di questo genere di documenti, sul finire del 2019, è stato stimato in oltre ottanta[223] – possono citarsi:

- per quanto riguarda le dichiarazioni diretta emanazione dell'assemblea di organi internazionali: le menzionate *Recommendation of the Council of OECD on Artificial Intelligence*[224], del 22 maggio 2019, e *Recommendation on the Ethics of Artificial Intelligence* dell'UNESCO[225], del 24 novembre 2021. Pur con rilevanti differenze, entrambi i documenti pongono l'accento sul rispetto dei diritti umani fondamentali da parte

[219] Cfr. *AI Strategy 2019 – Ai for everyone: People, Industries Regions and Governments*, 11 giugno 2021, disponibile in versione inglese a cura dello stesso governo giapponese, https://www8.cao.go.jp/cstp/ai/aistratagy2019en.pdf (7 novembre 2021).

[220] Il piano è disponibile in versione inglese all'indirizzo: https://k-erc.eu/wp-content/uploads/2017/12/Master-Plan-for-the-intelligent-information-society.pdf (7 novembre 2021). Il quadro è stato poi aggiornato nel 2019 con la pubblicazione di una nuova *AI Strategy*: https://bit.ly/31IqX5g (7 novembre 2021).

[221] Cfr. ASIA PACIFIC FOUNDATION OF CANADA, *Artificial intelligence policies in East Asia: an overview from the Canadian Perspective*, 2019, https://www.asiapacific.ca/sites/default/files/filefield/ai_report_2019.pdf (7 novembre 2021) p. 24 ss.

[222] *Ibidem*, p. 26.

[223] A. JOBIN, M. IENCA, E. VAYENA, *The global landscape of AI ethics guidelines*, in *Nature Machine Intelligence*, 1, 2019, p. 389-399, ad esempio, ne identificano 84.

[224] OECD, *Recommendation of the Council on Artificial Intelligence*, 22 maggio 2019, OECD/LEGAL/0449.

[225] UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, 24 novembre 2021, 41 C/73 (Annex).

dell'intelligenza artificiale, sui principi di trasparenza e comprensibilità dei risultati dei sistemi intelligenti e sulla permanenza di forme di controllo umano sul comportamento di quest'ultimi. Entrambi gli organismi internazionali, inoltre, evidenziano la necessità di orientare lo sviluppo dell'intelligenza artificiale verso il benessere collettivo e la sostenibilità ambientale.

- lo *Statement on Trade and Digital Economy* siglato nella riunione del G20 del 9 giugno 2019, e il relativo allegato contenente i *G20 AI Principles*[226]. Nella dichiarazione i leader dei paesi più sviluppati del mondo hanno affrontato, in generale, le problematiche connesse allo sviluppo tecnologico, concordando sulla necessità di promuovere lo sviluppo di una *human-centered artificial intelligence* [227] . I principi alla base di quest'ultima, indicati dal menzionato allegato, sono: sviluppo sostenibile, centralità dell'essere umano, equità, trasparenza e comprensibilità, sicurezza, possibilità di collocare la responsabilità per i sistemi intelligenti.
- le già prese in considerazione *Ethics Guidelines for trustworthyartificial intelligence*, elaborate dal Gruppo di Esperti d'Alto Livello nominato dalla Commissione Europea e pubblicate, nella versione finale, l'8 aprile 2019, le quali, allo stato dell'arte, sono il documento più completo in materia tra quelli riferibili, almeno indirettamente, a poteri pubblici nazionali o sovranazionali. Le linee guida prevedono che lo sviluppo, la distribuzione e l'utilizzo di tecnologie basate sull'intelligenza artificiale debbano conformarsi ai principi di rispetto dell'autonomia umana, prevenzione dei danni, equità ed esplicabilità, e indicano sette requisiti la cui esistenza dev'essere garantita per lo sviluppo di un'IA affidabile: intervento e sorveglianza umani, robustezza tecnica e sicurezza, riservatezza e governance dei dati, trasparenza, non discriminazione ed equità, benessere sociale e ambientale e collocamento della responsabilità.
- in riferimento ai contributi provenienti da iniziative promosse dal mondo della ricerca e da associazioni non governative: i 23 Principi di Asilomar sull'IA[228], frutto della *Asilomar Conference on Beneficial AI*tenutasi nel 2017 a Pacific Grove, California, organizzata dal *Future of Life Institute*, un think-tank specializzato nella promozione di uno sviluppo tecnologico armonico e orientato al progresso dell'essere umano[229]; il policy paper del 2019

[226] G20 MINISTERIAL STATEMENT ON TRADE AND DIGITAL ECONOMY, 9 giugno 2019, p. 3-4, https://trade.ec.europa.eu/doclib/docs/2019/june/tradoc_157920.pdf (7novembre 2021).
[227] Facendo, peraltro, riferimento ai principi elaborate dall'OCSE: «To foster public trust and confidence in AI technologies and fully realize their potential, we are committed to a human-centered approach to AI, guided by the G20 AI Principles drawn from the OECD Recommendation on AI», G20 MINISTERIAL STATEMENT ON TRADE AND DIGITAL ECONOMY cit., p. 3-4.
[228] ASILOMAR AI PRINCIPLES 2017, https://futureoflife.org/ai-principles/ (7 novembre 2021).
[229] Cfr. https://futureoflife.org/ (7 novembre 2021).

Understanding artificial intelligence ethics and safety: a guide for the responsible design and implementation of AI systems in the public sector, a cura del britannico Alan Turing Institute, che propone un'approfondita serie di concreti accorgimenti e protocolli di sicurezza al fine di minimizzare i rischi connessi allo sviluppo dell'IA[230]; i lavori del progetto AI4People, un forum di esperti provenienti dalla ricerca, dall'impresa e dalla società civile nato nel 2017 dalla collaborazione tra il Digital Ethics Lab dell'Università di Oxford e l'Atomium European Institute for Science, Media and Democracy[231], tra i cui risultati più rilevanti possono citarsi i paper *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*[232], pubblicato nel 2018, e *AI4Peoples 7 Global Frameworks* [233], del 2020, anch'essi orientati a fornire raccomandazioni concrete ai poteri pubblici e al mondo produttivo per lo sviluppo di un'intelligenza artificiale finalizzata al benessere collettivo, senza limitarsi a enumerazioni di principi.

- una menzione a parte, per il differente contesto politico e geografico di riferimento, meritano i Principi di Pechino sull'IA, stilati nel maggio 2019 da un gruppo di lavoro coordinato dal governo cinese e composto da esperti provenienti da diversi centri di ricerca e rappresentanti del mondo industriale, col coinvolgimento, tra le altre, di Baidu, Alibaba e Tencent [234]. Il documento rispecchia in larga misura quelli elaborati in Occidente, specialmente in riferimento ai principi di trasparenza, esplicabilità, collocamento della responsabilità e progresso orientato verso il benessere collettivo. Inoltre, posta la provenienza della lista di principi, è particolarmente significativo che prenda in considerazione l'impatto sui diritti individuali, pur con una formula che, dal punto di vista della nostra tradizione giuridica, non può non apparire tenue: *«Human privacy, dignity, freedom, autonomy, and rights should be sufficiently respected»*[235].

Al proliferare di documenti di soft-law e di piani strategici elaborati a livello nazionale e sovranazionale fa da contraltare l'assenza, allo stato dell'arte quasi totale, di strumenti di hard-law direttamente rivolti alla regolazione dell'intelligenza artificiale. L'unica eccezione di rilievo è

[230] D. LESLIE, *Understanding artificial intelligence ethics and safety: a guide for the responsible design and implementation of AI systems in the public sector*, The Alan Turing Institute, 2019https://doi.org/10.5281/zenodo.3240529 (23 novembre 2021).
[231]Cfr. https://www.eismd.eu/ai4people/ (7 novembre 2021).
[232]L. FLORIDI, J. COWLS, M. BELTRAMETTI, R. CHATILA, P. CHAZERAND, V. DIGNUM, C. LUETGE, R. MADELIN, U. PAGALLO, F. ROSSI, B. SCHAFER, P. VALCKE, E. VAYENA, *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, 2018, https://bit.ly/3pNOA4a (7novembre 2021).
[233] A.A.V.V., *AI4Peoples 7 AI Global Frameworks*, 2018, https://bit.ly/3rUxgNG (7 novembre 2021).
[234] È disponibile una traduzione non ufficiale in inglese dei principi di Pechino, a cura della rivista *Wired*: https://www.wired.com/beyond-the-beyond/2019/06/beijing-artificial-intelligence-principles/ (7 novembre 2021).
[235] La citazione proviene dalla traduzione indicata alla nota precedente.

rappresentata dalla già menzionata *Directive on Automated Decision-Making*, approvata a livello federale dal governo canadese nel 2019 e pienamente in vigore dal 1 aprile 2020[236]. La direttiva prevede un articolato sistema di garanzie che devono accompagnare lo sviluppo e l'utilizzo da parte della pubblica amministrazione di sistemi di decisione automatica, quali lo svolgimento di un *algorithmic impact assessment* e di un articolato insieme di test prima dell'implementazione, il diritto dell'interessato a una spiegazione della decisione, la possibilità di controllo e intervento umani sul sistema[237]. Oltre all'iniziativa canadese, si registrano proposte di regolazione –in Francia, nella scorsa legislatura, ne fu presentata addirittura una di rango costituzionale[238] - la cui approvazione verrà discussa nel prossimo futuro. La più rilevante, di cui subito si dirà, è senza dubbio la già nominata proposta di Regolamento sull'intelligenza artificiale presentata dalla Commissione Europea il 21 aprile 2021[239]. Infine, per completare lo scarno quadro della legislazione esistente in materia di intelligenza artificiale, deve menzionarsi la normativa in materia di trattamento e protezione dei dati personali, ormai presente in pressoché ogni paese industrializzato[240]. Pur abbracciando un oggetto di regolazione diverso e più ampio, infatti, le norme in materia di protezione dei dati personali, in particolare quelle di più recente elaborazione, hanno un impatto anche sulle tecnologie basate sull'intelligenza artificiale, posto il ruolo primario dei dati nello sviluppo di quest'ultime[241].

[236] GOVERNMENT OF CANADA, *Directive on Automated Decision-Making*, 2019, https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592 (7 novembre 2021).
[237] Cfr. in particolare il par. 6 della Direttiva.
[238] Si tratta della *Proposition de loi constitutionelle n. 2585 relative à la Charte de l'intelligence artificielle et des algorithmes* presentata il 15 gennaio 2020 dal deputato M. Pierre-Alain Raphan, che prevedeva l'approvazione, da parte dell'Assemblea nazionale, di una *Charte de l'intelligence artificielle et des algorithmes* e l'inserimento di un riferimento a quest'ultima nel Preambolo della Costituzione francese, mai discussa dall'*Assemblée nationale* nel corso della XV Legislatura (2017-2022).
[239] COMMISSIONE EUROPEA, *Proposta di Regolamento al Parlamento Europeo e al Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (Legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione*, 21 aprile 2021, COM(2021) 206 final. Per alcunicommenti v. C. CASONATO, B. MARCHETTI, *Prime osservazioni sulla Proposta di Regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *BioLaw Journal-Rivista di BioDiritto,* 3, 2021; N.A. SMUHA ET AL., *How the EU Can AchieveLegallyTrustworthy AI: A Response to the EuropeanCommission'sProposal for an Artificial Intelligence Act*, 2021, http://dx.doi.org/10.2139/ssrn.3899991; P. HACKER, *A legal framework for AI training data – from first principles to the Artificial Intelligence Act*, 2021, doi:10.1080/17579961.2021.1977219; A. LAVORGNA, G. SUFFIA, *La nuova proposta europea per regolamentare i Sistemi di Intelligenza Artificiale e la sua rilevanza nell'ambito della giustizia penale: un passo necessario, ma non sufficiente, nella giusta direzione*, in *Diritto Penale Contemporaneo*, 2, 2021, p. 88 ss.; F.C. LA VATTIATA, *Brevi note "a caldo" sulla recente Proposta di Regolamento UE in tema di intelligenza artificiale*, in *Diritto Penale e Uomo*, 6, 2021.
[240] Cfr. F. MOLNAR-GABOR, *Data Protection*, in *Max Planck Encyclopedia of Comparative Constitutional Law*, Oxford, 2016, https://ssrn.com/abstract=2883707 (7 novembre 2021).
[241] Tra i molti, si rimanda ancora a A. MANTELERO, *Artificial Intelligence and Data Protection cit.*; G. FINOCCHIARO, *Intelligenza artificiale e protezione dei dati personali cit.*

3. La proposta di Regolamento dell'Unione Europea sull'intelligenza artificiale

Il 21 aprile 2021 la Commissione Europea, seguendo l'iter previsto per la c.d. procedura legislativa ordinaria, ha reso pubblica una proposta di Regolamento indirizzata al Parlamento europeo e al Consiglio *«che stabilisce regole armonizzate sull'intelligenza artificiale (Legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione»*[242]. L'*Explenatory Memorandum* che introduce l'atto indica come basi giuridiche in prima battuta l'art. 114 TFUE, che legittima l'intervento dell'Unione per l'instaurazione e il funzionamento del mercato interno[243], e in seconda battuta l'art. 16 del medesimo Trattato, relativo alla protezione dei dati personali dei cittadini europei[244].

La proposta, in realtà, è il coronamento di un percorso di elaborazione normativa attorno all'IA lungo almeno tre anni, cominciato con la menzionata comunicazione della Commissione *L'intelligenza artificiale per l'Europa* del 2018 e proseguito con le *Linee Guida Etiche per un'IA affidabile* del Gruppo di Esperti di Alto Livello sull'IA, la relativa *Assessment List* e diversi altri documenti, a dimostrazione della profondità dell'interesse per il tema nel contesto unionale[245].

Il progetto normativo ha l'obiettivo dichiarato di affrontare i pericoli connessi ad alcune applicazioni dell'intelligenza artificiale, promuovendo, allo stesso tempo, lo sviluppo e la diffusione di tale tecnologia senza frustrare il mercato[246]. A tal fine, la Commissione ha scelto un approccio

[242] Per gli estremi cfr. poco sopra, n. 239.

[243] Il primo paragrafo dell'articolo recita: «Salvo che i Trattati non dispongano diversamente, si applicano le disposizioni seguenti per la realizzazione degli obiettivi dell'articolo 26. Il Parlamento europeo e il Consiglio, deliberando secondo la procedura legislativa ordinaria e previa consultazione del Comitato economico e sociale, adottano le misure relative al ravvicinamento delle disposizioni legislative, regolamentari ed amministrative degli Stati membri che hanno per oggetto l'instaurazione e il funzionamento del mercato interno».

[244] L'articolo recita: «1. Ogni persona ha diritto alla protezione dei dati di carattere personale che la riguardano. 2. Il Parlamento europeo e il Consiglio, deliberando secondo la procedura legislativa ordinaria, stabiliscono le norme relative alla protezione delle persone fisiche con riguardo al trattamento dei dati di carattere personale da parte delle istituzioni, degli organi e degli organismi dell'Unione, nonché da parte degli Stati membri nell'esercizio di attività che rientrano nel campo di applicazione del diritto dell'Unione, e le norme relative alla libera circolazione di tali dati. Il rispetto di tali norme è soggetto al controllo di autorità indipendenti. 3. Le norme adottate sulla base del presente articolo fano salve le norme specifiche di cui all'articolo 39 del trattato sull'Unione europea».

[245] L'*Explanatory memorandum*(p. 1-3) menziona, tra gli altri, il *Libro bianco sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia* della Commissione Europea (COM(2020) 65 final); la Comunicazione della Commissione *Plasmare il futuro digitale dell'Europa* (COM(2020) 67 final); la Risoluzione del Parlamento europeo del 20 ottobre 2020 recante *raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale*, *della robotica e delle tecnologie correlate*. 2020/2012(INL); la Risoluzione del Parlamento europeo del 20 ottobre 2020 *recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale*, 2020/2014(INL); la Risoluzione del Parlamento europeo del 20 ottobre 2020 *sui diritti di proprietà intellettuale per lo sviluppo di tecnologie di intelligenza artificiale*,2020/2015(INI); il Progetto di relazione del Parlamento europeo sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale, 2020/2016(INI); il Progetto di relazione del Parlamento europeo sull'intelligenza artificiale nell'istruzione, nella cultura e nel settore audiovisivo, 2020/2017(INI).

[246]L'*Explanatory memorandum*(p. 2) identifica quattro obiettivi: «assicurare che i sistemi di IA immessi sul mercato dell'Unione e utilizzati siano sicuri e rispettino la normativa vigente in materia di diritti fondamentali e i valori dell'Unione; assicurare la certezza del diritto per facilitare gli investimenti e l'innovazione nell'intelligenza artificiale; migliorare la governance e l'applicazione effettiva della normativa esistente in materia di diritti fondamentali e requisiti di sicurezza applicabili ai sistemi di IA; facilitare lo sviluppo di un mercato unico per applicazioni di IA lecite, sicure e affidabili nonché prevenire la frammentazione del mercato».

basato sulla gestione del rischio, mutuando un'impostazione già adottata, in altre forme, in precedenti atti normativi, a cominciare dal GDPR[247]. La proposta suddivide le applicazioni dell'intelligenza artificiale in quattro classi di rischio, sottoponendo ciascuna di esse a un differente regime normativo. Sono individuate, in primo luogo, talune applicazioni dell'intelligenza artificiale vietate nel contesto dell'Unione, per la potenziale lesività della dignità della persona e di un'ampia gamma di diritti individuali. Si tratta, ad esempio, di tecnologie volte a influenzare il comportamento di una persona al fine di orientarlo in senso dannoso verso sé stessa o gli altri (art. 5 par. 1 lett. a), di sistemi finalizzati a sfruttare specifiche vulnerabilità della persona quali disabilità o minore età (art. 5 par. 1 lett. b), di sistemi di valutazione del credito sociale che facciano discendere conseguenze pregiudizievoli per i soggetti coinvolti in contesti non connessi a quello in cui i dati di partenza sono stati raccolti (art. 5 par. 1 lett. c). Un regime speciale è riservato all'uso di sistemi di identificazione biometrica in tempo reale, consentito unicamente al fine di ricercare vittime di reato e minori scomparsi, prevenire il compimento di attacchi terroristici o perseguire autori e sospettati dei gravi reati indicati all'art. 2 par. 2 della Decisione quadro 2002/58/GAI in materia di mandato d'arresto europeo (art. 5 par. 1 lett. d).In secondo luogo, un ampio elenco di applicazioni dell'intelligenza artificiale, individuato dall'art. 6 della Proposta, è considerato ad alto rischio. La norma delimita il perimetro della categoria attraverso un sistema di rinvii agli allegati del testo legislativo, risultando di non facile lettura. Sono applicazioni ad alto rischio, prima di tutto,quelle indicate all'allegato III, individuate per la delicatezza del contesto di utilizzo e delle finalità per cui sono impiegate. L'elenco comprende, ad esempio, i sistemi utilizzati nell'attività creditizia, per la gestione dei flussi migratori, o per l'identificazione biometrica nei casi, appena visti, in cui il precedente art. 5 la permette. Rientrano nella categoria, inoltre, le tecnologie basate sull'intelligenza artificiale che costituiscano prodotti, o componenti di sicurezza di prodotti, disciplinati da un'ampio elenco di testi legislativi europei di armonizzazione, previsto all'allegato II, e per i quali tali norme prevedano una valutazione di conformità da parte di terzi. Le norme indicate in tale allegato II corrispondono in larga parte al c.d. *new legislative framework*, un pacchetto di misure adottato, a partire dal 2008, al fine di uniformare e migliorare gli standard di conformità richiesti per la commercializzazione nel mercato unico europeo di determinati prodotti[248]. I sistemi ad alto rischio dovranno rispettare stringenti requisiti in materia di qualità dei dataset, trasparenza, comprensibilità dei risultati e possibilità di controllo e intervento umani sul loro funzionamento. A tal fine, in capo ai fornitori di tali tecnologie sono posti specifici obblighi afferenti all'elaborazione di un sistema

[247] Cfr. tra i molti R. GELLERT, *Understanding the notion of risk in the General Data Protection Regulation*, in *Computer Law & Security Review*, 34, 2, 2018, p. 279-288.

[248] Si rinvia alle informazioni in proposito rese disponibili sul sito della Commissione Europea: https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en (23 ottobre 2022).

procedurale e documentale di *management* del rischio, e la loro immissione sul mercato dovrà essere preceduta da una valutazione di conformità ai requisiti richiesti dalla Proposta (art. 43) cui segue, in caso di esito positivo, l'apposizione sul prodotto del marchio CE (art. 49). Tali requisiti, invece, non sono considerati vincolanti per la commercializzazione dei sistemi di IA diversi da quelli ad alto rischio, che formano la terza classe di pericolosità considerata dal regolamento. La messa sul mercato di queste tecnologie, nelle intenzioni della Proposta, sarà lasciata in larga parte all'autoregolazione, tanto che il testo prevede l'invito esplicito agli operatori del settore a stipulare codici di condotta, allo scopo di aderire, adattandoli ove necessario, ai requisiti previsti per la diffusione e l'utilizzo di sistemi ad alto rischio (art. 69). Infine, la Proposta della Commissione prevede oneri specifici di trasparenza per alcuni particolari sistemi di IA, a prescindere dalla classe di rischio. Si tratta delle tecnologie usate per l'interazione diretta con l'essere umano, per rilevare emozioni, porre in essere attività di profilazione basate su dati biometrici o generare e manipolare contenuti estremamente realistici (c.d. *deep fake*). In tali casi, la Proposta di Regolamento prevede l'obbligo di informare i soggetti coinvolti dell'interazione con un agente intelligente, dell'utilizzo di quest'ultimo con finalità di profilazione emotiva o biometrica o della natura artificiale del contenuto generato dall'IA (art. 52).

Gli adempimenti previsti dalla Proposta di Regolamento sono accompagnati da un ventaglio di sanzioni, anche molto severe, per il caso di loro mancato rispetto. L'atto (art. 71) determina unicamente gli importi per le violazioni più significative, oscillanti da un minimo di 10.000.000 di Euro, o del 2% per cento del fatturato mondiale annuo dell'esercizio precedente, se superiore, a un massimo di 30.000.000 di Euro o del 6% del fatturato annuo. La sanzione massima è prevista per l'utilizzo delle applicazioni dell'intelligenza artificiale bandite dall'articolo 5 e per il mancato rispetto dei criteri di integrità dei dataset previsti dall'art. 10. La definizione delle sanzioni per le violazioni non considerate dalla Proposta, invece, è demandata a ciascuno stato membro.

Presentata, come già detto, il 21 aprile 2021, la Proposta è stata sottoposta ad un procedimento di consultazione pubblica della durata di poco più di 3 mesi (26 aprile – 4 agosto) cui è seguito un periodo di ridiscussione del testo, sulla base degli input ricevuti, prima di procedere con l'iter legislativo sottoponendo l'atto alla prima lettura del Parlamento[249]. Sebbene non sia possibile fare previsioni chiare sulla durata dell'intero procedimento legislativo – né, in realtà, sul suo stesso esito positivo – pare realistico ipotizzare che la Proposta possa trovare definitiva approvazione nel corso del 2023 o nella prima metà del 2024[250]. L'effettiva entrata in vigore potrebbe, in ogni caso, essere

[249] I risultati della consultazione sono disponibili sul sito della Commissione: https://bit.ly/3DMJAlF (7 novembre 2021).

[250] I tempi medi di completamento della procedura legislativa ordinaria oscillano tra i 18 e i 40 mesi, a seconda che l'atto sia approvato in prima o seconda lettura (solo un'esigua minoranza di atti è, invece, adottata in terza lettura). Cfr.

ulteriormente posticipata al fine di dare agli operatori del settore la possibilità di adeguare in anticipo la loro condotta, com'è avvenuto, ad esempio, nel caso del GDPR.

4. Lo stato dell'arte del diritto dell'intelligenza artificiale e alcune tendenze generali del suo sviluppo

In sintesi, il diritto dell'intelligenza artificiale, allo stato dell'arte, è formato più da intenzioni, proposte e prospettive di regolazione che da norme vincolanti. Infatti, alla vista varietà di piani strategici di sviluppo elaborati a livello nazionale e sovranazionale si accompagnano numerose dichiarazioni di principio, non vincolanti né azionabili in giudizio, e poche discipline di hard-law già in vigore. Parallelamente, sono in fase di studio e discussione alcune proposte per l'adozione di una normativa specifica per l'intelligenza artificiale, la più importante delle quali pare, come già detto, l'*Artificial Intelligence Act* europeo. Per quanto il quadro descritto nei paragrafi precedenti sia in rapida evoluzione, è possibile individuare due tendenze generali – oltre alla già menzionata prevalenza della soft-law sull'hard-law in questo momento storico - che caratterizzano i tentativi di regolazione, intesa in senso ampio, dell'intelligenza artificiale da parte dei poteri pubblici.

In primo luogo, la volontà di incanalare lo sviluppo tecnologico verso un'intelligenza artificiale centrata sull'essere umano, senza per questo ingessare il mercato. Emerge la consapevolezza, presente pressoché in tutti gli ordinamenti presi in esame, della necessità di governare i rischi connessi ad alcuni sistemi intelligenti, abbinata al timore di perdere posizioni nella corsa allo sviluppo tecnologico con forme di regolazione eccessivamente onerose. Elaborare strumenti normativi all'altezza di entrambi gli obiettivi sarà, probabilmente, una delle sfide centrali della regolazione dell'intelligenza artificiale.

In secondo luogo, la preferenza per testi normativi, sia di soft-law che di hard-law, d'ampio respiro, volti a definire regole applicabili all'intero insieme delle tecnologie basate sull'intelligenza artificiale e non a disciplinarne solo determinati settori e applicazioni[251]. Si tratta, probabilmente, di un approccio dovuto alla radicale mancanza di normativa sul tema, che rende necessario definire,

---

M. MCGUINNESS, E. GEBHARDT, P. TELICKA, C. WILKSTROM, *Activity report. Developments and trends of the ordinary legislative procedure – 1 July 2014 – 1 July 2019 (8th parliamentary term)*, 2019, https://www.europarl.europa.eu/cmsdata/198038/activity-report-2014-2019_en.pdf (10 dicembre 2021).

[251] Fa eccezione un testo normativo presentato nell'agosto del 2021 dal Dipartimento per l'Amministrazione del Cyberspazio della Repubblica Popolare Cinese ed entrato in vigore il primo aprile 2022, l'*Internet Information Service AlgorithmicRecommendation Management Provisions*,volto a regolare un campo di applicazione dell'intelligenza artificiale estremamente limitato e circoscritto, la raccomandazione agli utenti di servizi online di contenuti attinenti all'informazione, e che si inserisce in un ordinamento in cui, allo stato dell'arte, non si rinvengono norme di *hard-law* dirette a disciplinare l'intelligenza artificiale nel suo insieme. Una traduzione non ufficiale a cura dell'Università di Stanford è disponibile al link: https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-opinon-seeking-draft/ (10 dicembre 2021); per un commento cfr. L. M. LAVENUE, J. M. MYLES, A.N. SCHNEIDER, *Evaluating China's New 'Internet Information Service Algorithmic Recommendation Management' Regulations*, in *Finnegan*, 21 aprile 2022, https://bit.ly/3gtaPvr (23 ottobre 2022).

prima di tutto, un corpo di regole e principi generali. È, forse, anche quello più adatto ad evitare il menzionato rischio di ingessare il mercato con una regolazione eccessiva, anche se si espone al pericolo, proprio di ogni disciplina di rango generale, di risultare scarsamente adatto a talune applicazioni concrete.

PARTE II
I diritti fondamentali di fronte all'intelligenza artificiale

Nuove sfide per "vecchi diritti". Il principio personalista nell'era dell'intelligenza artificiale: l'evoluzione dei diritti a protezione della sfera dell'identità

1. Il diritto all'identità personale: origini, contenuto, e primo impatto con la rivoluzione digitale

La parola "identità" non è presente nella Costituzione italiana. La memoria della precedente esperienza autoritaria ha probabilmente dissuaso i Costituenti dall'inserire qualunque riferimento a una possibile identità collettiva (anche se l'idea in parte rimane nel concetto di nazione, più volte richiamato dalla Carta)[252]. Il termine, però, non compare nemmeno associato ad attributi come "individuale" o "personale", al fine di dare copertura costituzionale espressa a una specifica prerogativa del singolo. Non è un caso, allora, che i confini della tutela dell'identità siano stati protagonisti, nel nostro ordinamento, di un dibattito dottrinale e giurisprudenziale particolarmente vivace ed esteso nel tempo[253].

[252] Propone questa ricostruzione A. MORELLI, *Persona e identità personale*, in *BioLaw Journal – Rivista di BioDiritto*, Special Issue 2, 2019, p. 45-47. Il riferimento a identità collettive è presente, invece, nel Trattato sull'Unione Europea (art. 4 par. 2), i cui redattori non si trovavano di fronte a un immediato precedente totalitario e vedevano nel dialogo con le identità nazionali uno degli elementi fondamentali – e più complessi - della costruzione europea: «L'Unione rispetta l'uguaglianza degli Stati membri davanti ai Trattati e la loro identità nazionale, insita nella loro struttura fondamentale, politica e costituzionale, compreso il sistema delle autonomie locali e regionali». Sulla declinazione del concetto di nazione nella Costituzione repubblicana, cfr. *ex multis* V. CRISAFULLI, D. NOCILLA, *Nazione*, in *Enciclopedia del diritto*, XXVII, 1977, p. 805 ss.

[253] In letteratura, possono richiamarsi, in via generale e senz'animo di completezza, le ricostruzioni di G. BAVETTA, *Identità (diritto alla)* in *Enciclopedia del diritto*, XIX, 1970, p. 953 ss.; V. ZENO-ZENCOVICH, *Onore*, *reputazione e identità personale*, in G. ALPA, M. BESSONE (A CURA DI), *La responsabilità civile*, Torino, III, 1987 e*Identità personale*, in *Digesto delle discipline privatistiche*, IX, 1993, p. 294 ss.; A. CERRI, *Identità personale*, in *Enciclopedia giuridica*, agg. IV, Roma, 1995; G. PINO, *Il diritto all'identità personale. Interpretazione costituzionale e creatività giurisprudenziale*, Bologna, 2003; *Il diritto all'identità personale ieri e oggi. Informazioni, mercato, dati personali*, in R. PANETTA, *Libera circolazione e protezione dei dati personali*, Milano, 2006, p. 257 ss. e *Identità personale*, in S. RODOTÀ, M. TALLACCHINI (A CURA DI), *Ambito e fonti del biodiritto*, in S. RODOTÀ, P. ZATTI (DIRETTO DA), *Trattato di biodiritto*, Milano, 2010, p. 297 ss.; L. TRUCCO, *Introduzione allo studio dell'identità individuale nell'ordinamento costituzionale italiano*, Torino, 2004; E. RAFFIOTTA, *Appunti in materia di diritto all'identità personale*, in *www.forumcostituzionale.it*, 26 gennaio 2010, https://bit.ly/3sEShL5 (14 gennaio 2022); G. FINOCCHIARO, *Identità personale (diritto alla)*, in *Digesto delle discipine privatistiche*, aggiornamento 2010, p. 721-738; A. MORELLI, *Persona e identità personale cit.*

La questione si riconnette, in primo luogo, all'interpretazione data da dottrina e giurisprudenza alla categoria dei c.d. diritti della personalità nei primi decenni della storia repubblicana, ai quali, dal punto di vista civilistico, la tutela dell'identità va ascritta [254]. Per lungo tempo, infatti, l'interpretazione restrittiva dell'area di risarcibilità del danno non patrimoniale da parte delle Corti, limitata essenzialmente al danno da reato, hafatto dubitare che i diritti della personalità fossero veri e propri diritti soggettivi azionabili in giudizio[255]. Per le stesse ragioni, inoltre, appariva incerta la possibilità di riconoscere tutela a situazioni diverse rispetto alle ipotesi puntuali previste dal diritto positivo[256], all'epoca limitate alle norme in materia di atti di disposizione del corpo, diritto al nome e protezione dell'immagine previste agli artt. 6-10 del Codice civile, ai diritti riguardanti la proprietà intellettuale riconosciuti dalla L. n. 633 del 1941 [257] e al diritto di rettifica garantito dall'art. 8 della Legge sulla stampa[258]. Ad ogni modo,dottrina e giurisprudenza hanno superato queste incertezze almeno a partire dagli anni '70[259]. Infatti, il riconoscimento nell'ordinamento italiano di un pieno diritto all'identità personale, inteso come corretta rappresentazione da parte di altri della propria persona, comprensiva della sfera relazionale, sociale ed ideologica, azionabile e risarcibile in giudizio pur in assenza di espressa previsione normativa, si fa generalmente risalire a un'ordinanza del Pretore di Roma datata 6 maggio 1974[260].Il caso riguardava l'utilizzo, in un manifesto di propaganda antidivorzista, dell'immagine di un uomo e una donna, le cui convinzioni riguardanti il referendum che si sarebbe tenuto di lì a poco erano però di segno opposto e che non erano, in ogni caso, nemmeno uniti in matrimonio. Lo stesso giudice, appena un giorno dopo,

[254]Sul tema, si rimanda, in generale e tra i molti, a A. DE CUPIS, *I diritti della personalità*, Milano, 1982; P. RESCIGNO, *Personalità (diritti della)*, in *Enciclopedia giuridica*, XXIV, 1990; V. ZENO-ZENCOVICH, *Personalità (diritti della)*, in *Digesto delle discipline privatistiche*, XIII, 1995; D. MESSINETTI, *Personalità (diritti della)*, in *Enciclopedia del diritto*, XXXIII, 1983, p. 355; G. MARINI, *La giuridificazione della persona. Ideologie e tecniche dei diritti della personalità*, in *Rivista di diritto civile*, I, 2006, p. 359 ss.; R. PARDOLESI, *Diritti della personalità*, in AIDA, 2005, p. 3 ss; G. RESTA, *Autonomia privata e diritti della personalità*, Napoli, 2005; *Diritti della personalità: problemi e prospettive*, in *Diritto dell'informazione e dell'informatica*, 2007, p. 1043; G. ALPA, G. RESTA, *Le persone e la famiglia. 1. Le persone fisiche e i diritti della personalità*, in R. SACCO (DIRETTO DA), *Trattato di diritto civile*, Torino, 2019, p. 145 ss.; R. CASO, *La società della mercificazione e della sorveglianza: dalla persona ai dati*, Milano, 2021, p. 99-120.

[255]Sul punto cfr. F. SANTORO PASSARELLI, *Dottrine generali del diritto civile*, Napoli, 1966, p. 50 ss. e i menzionati D. MESSINETTI, *Personalità cit.*; P. RESCIGNO, *Personalità cit.*; G. MARINI, *La giuridificazione della persona cit.*; R. CASO, *La società della mercificazione e della sorveglianza cit.*, p. 99-108.

[256]Cfr. ancora, ad es., la ricostruzione di P. RESCIGNO, *Personalità cit.*, p. 5 ss.

[257]Legge n. 633 del 22 aprile 1941, *Protezione del diritto d'autore e di altri diritti connessi al suo esercizio*.

[258]Legge n. 47 dell'8 febbraio 1948, *Disposizioni sulla stampa.*

[259]Offre un sunto efficace di quanto accaduto V. ZENO-ZENCOVICH, *Identità personale cit.*, p. 294-295: «Se per un verso la paternità dell'espressione "identità personale" sembra doversi senz'altro attribuire al De Cupis (anche se non va dimenticata – per l'autorevolezza dello scrittore – la teoria di Ascarelli sulla paternità delle proprie azioni) la sua fortuna si manifesta solo un trentennio più tardi a seguito di una serie di convegni e seminari specificamente dedicati alla questione, nei quali l'autonomia della figura viene esaminata sotto i più diversi aspetti (non solo civilistici, ma anche costituzionali, penali, processuali) e la allora scarsa giurisprudenza viene sottoposta ad una vera e propria dissezione. Nel giro di un breve volgere di anni – la prima metà degli '80 – l'identità personale approda alla Corte di Cassazione, che, con la sentenza 22-6-1985, n. 3769, ne sancisce la rilevanza». Cfr. inoltre l'approfondita ricostruzione di R. CASO, *La società della mercificazione e della sorveglianza cit.*, p. 190 ss.

[260]Pretura di Roma, 6 maggio 1974, in *Foro italiano*, 1974, I, p. 1806; per un commento si veda la nota di A. D'ANGELO, *Lesione dell'identità personale e tutela riparatoria*, in *Giurisprudenza italiana*, 6, 1975, p. 515-518.

avrebbe emanato un provvedimento analogo, accogliendo le istanze del leader comunista Palmiro Togliatti, una cui dichiarazione era stata maliziosamente estrapolata dal contesto di riferimento e pubblicata, in un differente manifesto, dagli stessi comitati antidivorzisti, al fine di suggerire un'inesistente contrarietà al divorzio dell'autore[261]. Il suggello definitivo da parte della Corte di Cassazione giungerà, invece, nel 1985, con la sentenza finale del noto "caso Veronesi", in cui i giudici di legittimità riconobbero le ragioni dell'oncologo Umberto Veronesi, da anni impegnato nella lotta al tabagismo e le cui dichiarazioni erano state travisate, a scopi pubblicitari, da un marchio di sigarette *light*[262]. La pronuncia, infatti, riconobbe esplicitamente l'esistenza di un diritto soggettivo a vedere rispettato, da parte dei terzi, il proprio «modo di essere nella realtà sociale»[263] al fine di «svolgere integralmente la propria personalità individuale»[264].

Se l'esistenza di un diritto soggettivo all'identità, azionabile in giudizio, non fu mai più messa in discussione da dottrina e giurisprudenza a partire dal 1985, lo stesso non può dirsi dell'inquadramento costituzionale di tale diritto[265]. La sentenza sul caso Veronesi, infatti, radicava la tutela dell'identità in un'interpretazione analogica dell'art. 7 c.c., in materia di diritto al nome, e nell'art. 2 della Costituzione, interpretandolo, facendo riferimento a una nota ricostruzione teorica, come una clausola aperta, idonea a fornire copertura costituzionale anche a interessi non tipizzati nella Carta. Allo stesso tempo, però, non si era spinta al punto di includere la garanzia dell'identità personale tra i diritti costituzionalmente garantiti, specificando che essi «erano soltanto quelli specificamente previsti dalle successive norme della Costituzione»[266]. L'interpretazione portava a conseguenze di certo incoerenti, puntualmente evidenziate in dottrina: si pensi solo a come la protezione dell'identità personale e degli altri interessi che concorrono in vario modo a definirla – come la riservatezza, i cui rapporti con l'identità saranno indagati nei prossimi paragrafi, o l'onore – interferisca con una prerogativa di sicuro rango costituzionale come la libera manifestazione del

[261]Pretura di Roma, 7 maggio 1974, in *Foro italiano*, 1974, I, p. 3227.
[262]Cass. civ., sez. I, sent. 22 giugno 1985 n. 3769, in *Foro italiano*, I, 1985, p. 2211; si vedano le note di F. MACIOCE, *L'identità personale in Cassazione: un punto d'arrivo e un punto di partenza* e M. DOGLIOTTI, *Il diritto all'identità personale approda in Cassazione*, in *Giustizia civile*, 1, 1985, p. 3049 ss.
[263]Cass. civ., sez. I, sent. 22 giugno 1985 n. 3769 cit., p. 2216.
[264]*Ibidem.*
[265]Sul punto confronta, in particolare, i già citati G. PINO, *Il diritto all'identità personale cit.*, p. 80 ss. e*Il diritto all'identità personale ieri e oggi cit.*, p. 261 ss.; CASO, *La società della mercificazione e della sorveglianza cit.*, p. 190 ss; G. ALPA, G. RESTA, *Le persone e la famiglia. 1.Le persone fisiche e i diritti della personalità*, p. 319 ss.
[266]La sentenza, infatti, chiariva: «L'identità personale integra un bene essenziale e fondamentale della persona, quello di vedersi rispettato dai terzi il suo modo di essere nella realtà sociale, ossia di vedersi garantita la libertà di svolgere integralmente la propria personalità individuale, sia nella comunità generale che nelle singole comunità particolari. Essa è tutelata nella forma del diritto soggettivo, nel quadro dei diritti della personalità, con strumenti tipici del diritto privato. Pur riconducendosi all'art. 2 Cost., il diritto soggettivo dell'identità personale non si inserisce fra i diritti costituzionalmente garantiti, essendo tali soltanto quelli specificamente previsti dalle successive norme della Costituzione. La sua regolamentazione va dedotta, per analogia, dalla disciplina prevista per il diritto al nome (art. 7 c.c.), essendo tale figura la più affine al diritto all'identità personale.», cfr. Cass. civ., sez. I, sent. 22 giugno 1985 n. 3769 cit., p. 2216.

pensiero[267]. Se il diritto all'identità personale è idoneo a limitare l'esercizio della libertà d'espressione, come nei casi citati, deve avere lo stesso rango di quest'ultima, e deve includersi tra i diritti costituzionalmente garantiti senza ulteriori distinguo. Questa impostazione è stata accolta dapprima dalla Corte Costituzionale, nella sentenza n. 13 del 1994, in cui si legge a chiare lettere: «è certamente vero che tra i diritti che formano il patrimonio irretrattabile della persona umana l'art. 2 della Costituzione riconosce e garantisce anche il diritto all'identità personale»[268]. Poco dopo, la stessaSuprema Corte è tornata sui propri passi, superando le ambiguitàdella pronuncia del 1985. I Giudici di Legittimità, nella sentenza n. 978 del 1996 (relativa al noto caso Tabocchini - Re Cecconi) hanno affermato esplicitamente che il diritto a definire liberamente la propria identità e vederla rispettata dai terzi ha pieno rango costituzionale, e, anzi, rappresenta una delle primarie vie di realizzazione del principio personalista[269]. Solo vedendo garantita dall'ordinamento la libertà di essere ed evolvere sé stessi senza indebite interferenze da parte di terzi, infatti, è possibile perseguire genuinamente, dal punto di vista individuale, il fine del «pieno sviluppo della persona umana»[270] che innerva l'intero ordinamento. Questa presa di posizione si accompagna, nella menzionata sentenza e nella dottrina civilistica oggi prevalente, all'adesione alla c.d. teoria monistica dei diritti della personalità, che vede la categoria come un unico diritto soggettivo allo svolgimento della propria personalità, che si manifesta, a seconda del caso concreto, in forme differenti, talune esplicitate dal Legislatore e altre ricavabili per analogia e dai principi generali dell'ordinamento[271]. Risultano definitivamente superate, in tal modo, le risalenti incertezze riguardo all'estensibilità della categoria.

---

[267] Cfr. in particolare G. PINO, *Teoria e pratica del bilanciamento: tra libertà di manifestazione del pensiero e tutela dell'identità personale*, in *Danno e responsabilità*, 6, 2003, p. 577-586e*Il diritto all'identità personale ieri e oggi cit.*, p. 262 ss.;A. BEVERE, A. CERRI, *Il diritto di informazione e i diritti della persona*, Milano, 1995, p. 154 ss.

[268]Corte cost., sent. n. 13 del 3 febbraio 1994 (ud. 24 gennaio 1994); per un commento si veda la nota di A. PACE, *Nome, soggettività giuridica e identità personale*, in *Giurisprudenza costituzionale*, 1, 1994, p.103-105.

[269] «individuare con maggiore risolutezza il correlativo fondamento giuridico [del diritto all'identità, *ndr*], ancorandolo direttamente all'art. 2 Cost. inteso tale precetto nella sua più ampia dimensione e suscettibile, per ciò appunto, di apprestare copertura costituzionale ai nuovi valori emergenti della personalità in correlazione anche all'obiettivo primario di tutela del "pieno sviluppo della persona umana", di cui al successivo art. 3 cpv.», Cass. civ., sez. I, sent. 7 febbraio 1996 n. 978, in *Foro italiano*, I, 1985, p. 2211; per un commento v. la nota di A. D'ADDA, *La Corte di Cassazione riafferma il proprio orientamento in tema di diritto all'identità personale*, in *Responsabilità civile e previdenza*, 2-3, 1997, p. 474-481.

[270]La citazione, ovviamente, proviene dall'art. 3 c. 2 della Costituzione italiana.

[271]Fa chiarezza sul punto un passo della sentenza: «Quest'ultima puntualizzazione [relativa al valore unitario della persona umana, *ndr*], che presuppone l'adesione ad una concezione "monistica" dei diritti della personalità (da questa Corte, del resto, già sostanzialmente anticipata nella citata sent. 990 del 1963) aiuta anche a definire, senza perplessità, in termini di diritto soggettivo perfetto, la struttura della situazione giuridica considerata», Cass. civ., sez. I, sent. 7 febbraio 1996 n. 978 cit. Per una ricostruzione del dibattito tra teoria monista e pluralista cfr. *ex multis* G. PINO, *Teorie e dottrine dei diritti della personalità. Uno studio di meta-giurisprudenza analitica*, in *Materiali per una storia della cultura giuridica*, 1, 2003, p. 237-274; R. CASO, *La società della mercificazione e della sorveglianza cit.*, p. 108 ss e, in epoca più risalente, P. RESCIGNO, *Personalità cit.*, p. 5-7, che aderisce alla tesi pluralista, al tempo non chiaramente minoritaria.

Dal punto di vista comparato, l'elaborazione teorica avvenuta in Italia è stata preceduta da quelle della dottrina e della giurisprudenza tedesche e francesi. La categoria dei diritti della personalità, infatti, è frutto del lavoro della dottrina tedesca del XIX secolo, che ha posto le basi del concetto di diritto generale della personalità (*allgemeines Personlichkeitsrecht*) tuttora prevalente nella cultura giuridica della Germania[272]. Da tale diritto, il cui aggancio normativo è oggi individuato dagli interpreti direttamente nella Carta di Bonn, la giurisprudenza ha fatto discendere una grande varietà di interessi meritevoli di tutela, tra cui anche il diritto all'identità, nella varietà di significati nei quali è inteso nell'ordinamento italiano[273].Se il primato dell'elaborazione teorica appartiene alla dottrina tedesca, è stata la giurisprudenza francese, invece, a riconoscere protezione per prima ai diritti della personalità, nonostante l'assenza di norme in materia nell'impostazione originaria del *code civil*. Ciò si deve, soprattutto, all'assenza di limiti espressi alla risarcibilità del danno patrimoniale e alla presenza, già nel XIX secolo, di strumenti di tutela particolarmente rapidi, elastici ed incisivi (si pensi al meccanismo delle *astreintes*)[274]. Tali circostanze hanno permesso una ricca elaborazione pretoria e il riconoscimento in via giurisprudenziale di un *droit a l'identité personnelle*[275].

La tutela dell'identità individuale, intesa come libera definizione e sviluppo della propria personalità e protezione da indebite interferenze di terzi, ha trovato riconoscimento anche nei sistemi di *common law*, ai quali la categoria dei diritti della personalità è del tutto estranea. Infatti, l'ampio concetto di autodeterminazione insito nell'idea di *privacy*, patrimonio di quegli ordinamenti, ha permesso il riconoscimento di una protezione della sfera dell'identità analoga a quella dei paesi dell'Europa continentale[276]. Ciò è vero anche nell'ordinamento statunitense, in cui è

[272] Cfr. sul punto, tra i molti, G. PINO, *Il diritto all'identità personale ieri e oggi cit.*, p. 264 ss; R. CASO, *La società della mercificazione e della sorveglianza cit.*, p. 99 ss; G. RESTA, *Autonomia privata e diritti della personalità cit.*, p. 43 ss. e 104 ss. L'elaborazione teorica tedesca si deve principalmente agli studi di Otto von Gierke, Karl von Gareis e Joseph Kohler. Tra le loro opere più rilevanti per il campo dei diritti della personalità possono menzionarsi O. VON GIERKE, *Allgemeiner Teil und Personenrecht*, Lipsia, 1895; K. VON GAREIS, *Der Allgemeine Teil des Burgerlichen Gesetzbuchs*, Berlino, 1900; J. KOHLER, *Recht und Persönlichkeit in der Kultur der Gegenwart*, Stoccarda, 1914.

[273] V. ancora G. PINO, *Il diritto all'identità personale ieri e oggi cit.*, p. 264 ss. La copertura costituzionale del *allgemeines Personlichkeits recht* è individuata nell'art. 2 c. 1 della *Grundgesetz*, che riconosce a ogni individuo il diritto al libero sviluppo della propria personalità. Cfr. anche A. SOMMA, *I diritti della personalità e il diritto generale della personalità nell'ordinamento privatistico della Repubblica Federale Tedesca*, in *Rivista trimestrale di diritto e procedura civile*, 3, 1996, p. 805 ss.

[274] Cfr. in particolare R. CASO, *La società della mercificazione e della sorveglianza cit.*, p. 99 ss; G. RESTA, *Autonomia privata e diritti della personalità cit.*, p. 30 ss.; G. PINO, *Il diritto all'identità personale ieri e oggi cit.*, p. 266 ss.

[275] Alcune esemplificazioni della creatività della giurisprudenza francese in materia di diritti della personalità sono state raccolte da M. BESSONE, *Principi della tradizione e nuove direttive in materia di diritto all'immagine*, in *Foro italiano*, IV, 1974, p. 182-184. Cfr. anche A. LEPAGE, L. MARINO, *Droits de la personnalité*, in *Recueil Dalloz*, 39, 2007, p. 2771 ss.

[276] Affronta il tema delle forme di protezione dei diritti della personalità negli ordinamenti anglosassoni, e in particolare in quello statunitense, G. PINO, *Il diritto all'identità personale ieri e oggi cit.*, p. 268 ss. Sulle differenti concezioni di *privacy* nei vari ordinamenti e le peculiarità dei sistemi anglosassoni v., tra i molti, A. BALDASSARRE, *Privacy e costituzione: l'esperienza statunitense*, Roma, 1974; A. JONSSON CORNELL, *The right to privacy*, in *The Max Planck Encyclopedia of Comparative Constitutional Law*, Oxford, 2017, https://bit.ly/3I2Xtim (21 gennaio 2022).

pacifica la risarcibilità del *tort*di*false light in the public eye*, pur con alcuni distinguo, facilmente comprensibili se si tiene a mente la particolare attenzione per il *free speech*che caratterizza quella tradizione costituzionale[277]. Nel caso, infatti, l'erronea rappresentazione pubblica dell'identità riguardi *public figures* – e dunque persone già note, e non semplici privati cittadini, per i quali la tutela è sostanzialmente equivalente al modello europeo – la risarcibilità è possibile, anche dal punto di vista strettamente civilistico, nei soli casi in cui la condotta lesiva sia di intensità tale da integrare, allo stesso tempo, il delitto di diffamazione dolosa[278]. In ogni caso, non si può trascurare che, al netto di questi limiti, proprio la tutela particolarmente intensa della libertà individuale – anche dal punto di vista della libertà d'espressione – alla base dell'ordinamento statunitense fa sì che in quella tradizione giuridica sia presente una protezione particolarmente intensa della libera costruzione ed evoluzione della personalità[279].

Svolto questo primo sintetico inquadramento, deve evidenziarsi che, a partire dagli anni '90 del XX secolo, il tema della protezione dell'identità personale è stato rivoluzionato dalla diffusione capillare dei computer e dall'elaborazione di norme sul trattamento dei dati personali sempre più complesse[280]. Le possibilità di archiviazione e analisi di un volume inedito di informazioni sull'individuo garantite dagli elaboratori elettronici hanno iniziato a porre rischi e questioni senza precedenti per la sfera della personalità morale, circostanza che la migliore dottrina aveva evidenziato fin dagli anni '70[281]. Infatti, è evidente che ogni diritto attinente in senso ampio all'identità della persona, a cominciare dall'integrità della rappresentazione pubblica di quest'ultima o dalla possibilità di tenerne riservati determinati aspetti, riguarda in primo luogo il controllo sulla comunicazione e circolazioni di dati personali. In Italia, lo hanno riconosciutogli stessi testi normativi in materia, che hanno indicato la protezione dell'identità tra i propri fini espliciti, fornendo, così, per la prima volta, indiretto riconoscimento legislativo a tale diritto[282].

---

[277] Per la derivazione del *tort* di *false light in the public eye* dall'area della *privacy* nella concezione americana cfr. W. PROSSER, *Privacy*, in *California Law Review*, 48, 1960, p. 383-423; W. L. KEETON, W. PROSSER, *On the law of torts*, St. Paul, 1984, p. 849 ss.; D. L. ZIMMERMAN, *False Light Invasion of Privacy: The Light That Failed*, in *New York University Law Review*, 64, 1989, p. 364-453.

[278] È l'orientamento seguito dalla Corte Suprema almeno a partire dai casi *Time, Inc. v. Hill* (1967) e *Cantrell v. Forest City Publishing Co.* (1974). Cfr., nella dottrina italiana, A. GAMBARO, *Falsa luce agli occhi del pubblico (false light in the public eye)*, in *Riv. dir. civ.*, 1981, p. 84-135; M. L. RUFFINI GANDOLFI, *Il diritto all'identità personale di fronte alla Corte Suprema degli Stati Uniti (il tort di false light in the public eye)*, in *Riv. dir. ind.*, 1981, p. 237-280.

[279] Riguardo alla tutela particolarmente intensa del *free speech* nell'ordinamento statunitense, v. *ex multis* G. R. STONE, L. C. BOLLINGER, *The free speech century*, New York, 2019;F. ABRAMS, *The soul of the first amendment*, New Haven, 2017; Z. CHAFEE, *Free speech in the United States*, Cambridge (USA), 1941.

[280] Per una panoramica sull'emanazione di norme sul trattamento dei dati personali nei diversi ordinamenti, v. F. MOLNAR-GABOR, *Data protection*, in *Max Planck Encyclopedia of Comparative Consitutional Law*, Oxford, 2016.

[281] Il riferimento, in primo luogo è al lavoro di Stefano Rodotà. Cfr., in particolare, S. RODOTÀ, *Elaboratori elettronici e controllo sociale*, Bologna, 1973 e gli scritti raccolti in *Tecnologie e diritti*, Bologna, 1995.

[282] Sia l'art. 1 c. 1 della L. 675/1996 che l'art. 2 c. 1 del D. Lgs. 196/2003 indicavano la protezione dell'identità personale tra le finalità perseguite dai rispettivi corpi normativi in materia di trattamento dei dati personali, assieme al rispetto di diritti e libertà fondamentali, della dignità dell'interessato e della riservatezza.

La compenetrazione tra protezione dei dati personali e diritti della personalità è tale da aver portato, in dottrina, ad affermare che una cosa non possa più darsi senza l'altra, e che ogni controversia in materia di tutela dell'identità sia, oggi, allo stesso tempo una controversia in materia di trattamento di dati personali[283]. Che si voglia o meno aderire *in toto* a questa tesi, è indubbio che la normativa sui dati personali, concepita per la rivoluzione informatica, sia oggi preponderante per complessità e dettaglio e rappresenti il primo riferimento in materia. Ciò rimane vero anche per l'analisi delle nuove sfide alla sfera dell'identità poste dall'intelligenza artificiale, cui saranno dedicati i prossimi paragrafi.

2. L'evoluzione del diritto all'identità personale di fronte a certe applicazioni dell'intelligenza artificiale: cenni tecnici su intelligenza artificiale *data driven*, profilazione, *nudging* e sistemi di credito sociale

2.1 Dall'espressione dell'identità alla sua formazione: intelligenza artificiale e nuovi rischi per il libero sviluppo della personalità

L'elaborazione teorica e giurisprudenziale analizzata nel paragrafo precedente ha definito il diritto all'identità come diritto a una corretta espressione all'esterno della propria personalità, e, ancor di più, a una fedele rappresentazione di quest'ultima da parte di terzi. È un'impostazione che dà per scontato che la formazione di tale personalità, attraverso l'acquisizione di esperienze e conoscenze e la costruzione di rapporti sociali, sia libera. E, in effetti, negli stati costituzionali di diritto contemporanei una varietà di altri diritti, in gran parte esplicitati nelle Carte fondamentali, presidia la libertà del foro interno di ciascuno da indebite interferenze esterne, al fine di permettere quel «libero sviluppo della persona umana», la cui base è, appunto, la libertà di definire sé stessi. Ad esempio, le garanzie connesse alla libertà d'espressione, la libertà di accedere a un sistema di informazione largamente pluralista, l'assenza di orizzonti etici e religiosi imposti dallo stato, o la libertà di associazione e riunione garantiscono che la formazione di ogni individuo avvenga con una libertà da condizionamenti esterni inimmaginabile in altre epoche storiche. Questa intera costruzione teorica, però, si radica nella cultura giuridica liberale e democratica del XIX e del XX secolo, e, come tale, prende in considerazione la società prima dell'avvento del digitale[284]. Secondo

[283] È la tesi di G. PINO, *Il diritto all'identità personale ieri e oggi cit.*, p. 301: «è difficile negare che, adesso, qualsiasi violazione del diritto all'identità personale non può che postulare un illecito o non corretto trattamento di dati personali. Si scorra a piacimento la variegata e talvolta fantasiosa casistica portata all'attenzione delle corti, nel corso della più che ventennale esistenza giurisprudenziale del diritto all'identità personale, e ci si accorgerà che tutte quelle fattispecie sono adesso traducibili – direttamente e senza residui – in altrettante fattispecie di trattamento di dati personali. Detto più chiaramente: non è più possibile immaginare una violazione del diritto all'identità personale che non passi attraverso un trattamento illecito o non corretto di dati personali».

[284] Cfr., in generale, N. MATTEUCCI, *Organizzazione del potere e libertà. Storia del costituzionalismo moderno*, Torino, 1976 e *Breve storia del costituzionalismo* (1964),Brescia, 2010; G. VOLPE, *Il costituzionalismo del Novecento*, Roma-Bari, 2000; M. FIORAVANTI, *Costituzionalismo. La storia, le teorie, i testi*, Roma, 2018.

questa prospettiva, la libertà del singolodeve essere tutelata innanzitutto neiconfrontidell'autorità dello stato, che il ricordo dei totalitarismi che hanno preceduto molte costituzioni contemporanee fa sembrare un rischio immanente, da imbrigliare in un complesso sistema di pesi e contrappesi. Le possibili interferenze dei poteri pubblici nella vita dei singoli avevano, nella mente dei giuristi dell'epoca, il volto dell'aperta repressione, dell'uso sproporzionato della forza e della tortura. Per questa ragione, l'attività di polizia e le misure privative della libertà personale vengono sottoposte a rigidi presidi procedurali, spesso con norme di dettaglio inserite direttamente in Costituzione, come nel caso italiano[285]. A quel tempo, la conoscenza da parte dello stato delle attività private di un cittadino pareva possibile solo col dispiegamento di ingenti sistemi di spionaggio e sorveglianza, come avvenivanello stato autoritario novecentesco.

Come già accennato, la comparsa dei primi elaboratori elettronici, e le capacità di archiviazione di dati che ne derivarono, portarono ad individuare rischi inediti per i diritti individuali, cui si è cercatodi rispondere attraverso l'elaborazione di normative di dettaglio in materia di privacy, al fine di non rendere vane le garanzie del costituzionalismo appena richiamate. Lo scopo di questo paragrafo e di quello successivo sarà evidenziare come lo sviluppo delle tecniche di archiviazione e analisi dei dati avvenuto negli ultimi decenni, e in particolare lo sviluppo di alcune tecnologie basate sull'intelligenza artificiale,abbia rinnovato tali rischi. L'elaborazione di dati personali con gli algoritmi di più recente generazione, infatti, permette una profilazione estremamente accurata degli individui, e l'utilizzo delle informazioni così ricavate per una pluralità di finalità, tra cui quella di influenzarne il comportamento[286]. Ne derivano, come si dirà, sfide inedite per la sfera dei diritti, e in primo luogo per il diritto all'identità, che viene in rilievo non solo nella concezione di diritto a una libera espressione e a una corretta rappresentazione di sé stessi, ma in quella,logicamenteantecedente, di diritto a una libera e autentica definizione di tale identità, e alla possibilità di mutarla. Prima di addentrarsi in questo profilo è necessario, però, analizzare brevemente le tecnologie a cui si fa riferimento.

[285]L'ovvio riferimento è agli artt. 13 e 14 della Costituzione. Cfr. ad es. G. AMATO, *Art. 13* e *Art. 14* in G. BRANCA (A CURA DI), *Commentario della Costituzione*, II, Bologna-Roma, 1977.

[286]Tra i molti cfr., in via generale e da prospettive differenti, J. LEE, *Postcards from Planet Google*, The New York Times, 28 novembre 2002; H. R. VARIAN, *Beyond big data*, NABE Annual Meeting – San Francisco, 10 settembre 2013, https://bit.ly/3I6w7Ia (22 gennaio 2022); S. MILLS, *Into hyperspace: an analysis of hypernudges and personalized behavioural science*, 2019, https://doi.org/10.2139/ssrn.3420211 (22 gennaio 2022); S. BAROCAS, A. D. SELBST, *Big Data's Disparate Impact*, in *California Law Review*, 104, 2016, p. 671-732; K. YEUNG, *Why Worry about Decision-Making by Machine?*, in K.YEUNG, M. LODGE (A CURA DI), *Algorithmic Regulation*, Oxford, 2019 e *"Hypernudge": Big Data as a mode of regulation by design*, in *Information, communication and technology*, 20, 1, 2017, p. 118-136;M. VON OTTERLO, *A machine learning view on profiling*, in *Privacy, due process and the computational turn: the philosophy of law meets the phylosophy of technology*, New York, 2013, p. 41-65; S. ZUBOFF, *The age of surveillance capitalism: the fight for a human future at the new frontier of power*, New York, 2018.

2.2 L'intelligenza artificiale applicata all'analisi dei dati: profilazione e nuove forme di nudging

È stato estensivamente chiarito nella prima parte di questo lavoro che il termine "intelligenza artificiale" abbracciaunnumeroso e variegato insieme di applicazioni. Ognuna di esse, nonostante le rilevanti differenze, ha tra i propri elementi caratterizzanti la memorizzazione e l'analisi di informazioni sul contesto di riferimento, condotta in modo almeno parzialmente autonomo. A tale grande varietà corrisponde un numero altrettanto elevato di questioni riguardanti l'interazione con l'essere umano e, nella prospettiva giuridica adottata da questo lavoro, l'eventuale impatto sui diritti fondamentali. Intuitivamente, le applicazioni dell'intelligenza artificialeil cui "materiale di lavoro" è formato principalmente da dati personali pongono problemi specifici, in primo luogo, come detto, per la sfera dell'identità. Esse sono impiegate principalmente nei servizi del c.d. Internet 2.0, basati sull'interazione con l'utente e sulla raccolta dei dati personali che esso diffonde[287]. Tali sistemi orientano l'elaborazione dei dati a un determinato risultato, spesso consistente in un *output* rivolto all'utente nello stesso contesto dei servizi internet (si pensi, per limitarsi a un esempio molto comune, alla sottoposizione a quest'ultimo di pubblicità personalizzata). Deve evidenziarsi, però, che la rapida diffusione del c.d. Internet of Things rende possibili applicazioni di questo genere di tecnologie in numerosissimi ambiti della vita quotidiana[288], facendo perdere di senso la stessa distinzione tra "mondo di internet" (inteso, nella concezione comune ancora prevalente, come un'esperienza unicamente digitale, fruita attraverso uno schermo) e mondo fisico[289].

Una distinzione di massima piuttosto comune riguardo a queste tecnologie è quella tra intelligenza artificiale *model-based* e *data-driven*, parzialmente già vista nella prima parte. In via di estrema semplificazione, l'approccio *model-based* si fonda sulla programmazione di regole che il sistema, nel suo funzionamento, automatizza, come avveniva per i menzionati sistemi esperti[290]. L'approccio *data-driven*, invece, è alla base dell'apprendimento automatico, e si basa sull'osservazione di dati

[287] Il termine, nella versione inglese *Web 2.0*, è stato coniato da D. DINUCCI, *Fragmented future*, in *Print*, 32, 2019, p. 220-223 ed è diventato di uso comune a partire dalla prima edizione della *O'Rielly Web 2.0 Conference*, tenutasi a San Francisco nel 2004, https://bit.ly/3sqzOTv (28 gennaio 2021).

[288] Cfr. *ex multis* P. MAGRASSI, T. BERG, *A World of Smart Objects,* Gartner research report R-17-2243, 12 agosto2002;K. ASHTON, *That "Internet of Things" thing*, *RFID – Journal*, 22 giugno 2009, https://bit.ly/3pjWJOw (28 gennaio 2022); D. UCKELMANN, M. HARRISON, F. MICHAHELLES (A CURA DI), *Architecting the Internet of Things*, Berlin-Heidelberg, 2011.

[289]Sul tema si rimanda, in primo luogo, al lavoro del filosofo Luciano Floridi. Cfr. ad esempio L. FLORIDI, *Infosfera – filosofia ed etica dell'informazione*, Torino, 2009; *The fourthrevolution – How the Infosphereisreshaping human reality*, Oxford, 2014; L. FLORIDI (A CURA DI), *The Onlife Manifesto*, Cham, 2015. Di particolare interesse, inoltre, risulta una recente teorizzazione di L. VIOLANTE, *Diritto e potere nell'era digitale. Cybersociety, cybercommunity, cyberstate, cyberspace: tredici tesi*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2022, p. 145-153.

[290]Cfr., in generale, S. RUSSELL, P. NORVIG, *Artificial intelligence cit.*, p. 47 ss.; T. WEI, X. CHEN, X. LI, Q. ZHU, *Model-based and data-driven approaches for building automation and control*, in *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2019, doi:10.1145/3240765.3243485; S. FORMENTIN, K, VON HEUSDEN, A. KARIMI, *Model-based and data-driven model-reference control: a comparative analysis*, in *2013 European Control Conference (ECC)*, 2013, 10.23919/ECC.2013.6669388.

da parte dell'algoritmo, al fine di trovare autonomamente modelli e correlazioni, senza la programmazione delle regole d'inferenza per individuarli[291]. Nonostante le due modalità siano sempre più spesso integrate all'interno della medesima tecnologia, le applicazioni dell'intelligenza artificiale che elaborano dati personali raccolti in internet vanno ricondotte prevalentemente all'impostazione *data driven*. Esse, infatti, si basano sull'elaborazione di una mole sempre maggiore di dati personali, che permette la ricostruzione di analogie e corrispondenze sempre più complesse, spesso del tutto incomprensibili per l'osservatore umano. L'analisi dei dati così condotta rende possibile ricondurre l'utente i cui dati sono trattati in una determinata categoria di soggetti aventi caratteristiche comuni (c.d. *clustering*, o, nel linguaggio adottato dalla normativa europea sui dati personali,"profilazione")[292]. Il profilo individuato è costantemente aggiornato e perfezionato dalla raccolta di nuovi dati e può essere utilizzato per proporre all'utente l'acquisto di determinati prodotti o perindovinarne le opinioni politiche, l'orientamento sessuale, le convizioni religiose, e, in ultima analisi, anche i tratti della personalità più intimi e spiacevoli. L'ovvio vantaggio commerciale rappresentato da questi algoritmi, e il fatto che il loro funzionamento non possa che perfezionarsi al crescere della quantità di dati disponibili, spiega l'attuale struttura del mercato dei servizi internet: pochi grandi operatori privati che forniscono gratuitamente servizi fondamentali per il normale funzionamento della rete(come i motori di ricerca) o in cui gli utenti diffondono spontaneamente i propri dati personali (come i social network) e che vedono i propri immensi guadagni provenire dalla vendita di servizi pubblicitari di efficacia incredibilmente maggiore rispetto a quelli tradizionali[293]. Tutto ciò assieme alla diffusione di strumenti per la profilazione degli utenti (si pensi ai *cookies* con tali finalità) pressoché in ogni sito internet[294].

Le tecnologie in esame, dunque, hanno una finalità prima di tutto conoscitiva e predittiva: vengono analizzati dati personali di ogni genere sul soggetto interessato, che, aggregati con quelli di milioni di altri soggetti, permettono di ricostruire un profilo accurato e predire preferenze e interessi a breve e lungo termine. Si tratta di informazioni che possono essere impiegate per una pluralità di obiettivi, che vanno, come già detto, dal marketing alla propaganda politica o all'automazione di decisioni e

[291]Cfr. S. A. YABLONSKY, *Multidimensional data-driven artificial intelligence innovation*, in *Technology innovation management review*, 9, 12, 2019, p. 16-28; K. MANHART, *Artificial Intelligence Modelling: Data Driven and Theory Driven Approaches*, in K. G. TROITZSCH (A CURA DI), *Social Science Microsimulation*, Berlino, 1996, p. 416-431.

[292]L'art. 4 c. 1 n. 4 del Reg. UE n. 679 del 2016 (GDPR) così definisce "profilazione": «qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica».

[293]Cfr. per un commento D. SCHARFENBERG, *WhyFacebook and Google shouldpayyou for your data*, The Boston globe, 14 giugno 2018 e, in generale e tra molti, L. HJORTH, S. HINTON, *Understanding social media (2nd ed.)*, Londra, 2019.

[294]Cfr. A. CAHN, S. ALFED, P. BARFORD, S. MUTHUKRISHNAN, *An empirical study of web cookies*, in *WWW '16: Proceedings of the 25th International Conference on World Wide Web*, 2016, p. 891-901; J. PIERSON, R. HEYMAN, *Social media and cookies: challenges for online privacy*, in *Info*, 13, 6, 2011, p. 30-42.

valutazioni un tempo svolte con le sole conoscenze, sensibilità ed esperienza degli umani incaricati di esse. Tali tecnologie, però, stanno dimostrando di poter essere utilizzate non solo per conoscere e predire preferenze e comportamento degli individui, ma anche per influenzarli e modificarli, come è stato evidenziato da diversi studi condotti nell'ultimo decennio[295]. Da tale punto di vista, viene in gioco, quale antecedente necessario, il concetto di *nudge*.

Il termine *nudge* è stato reso popolare da un fortunato libro del 2008 dell'economista Richard Thaler e del giurista Cass Sunstein, *Nudge: improving decisions about health, wealth and happiness*[296]. Le tesi esposte nel testo riposano sui risultati ottenuti nei decenni precedenti nel campo dell'economia comportamentale e della psicologia cognitiva sulle reali modalità di decisione degli esseri umani, ben distanti dagli standard di razionalità alla base della teoria del consumatore neoclassica. In particolare, gli studi degli psicologi israeliani Daniel Kahneman e Amos Tversky e dello stesso Thaler hanno dimostrato che i processi decisionali sono normalmente viziati da una grande varietà di bias cognitivi[297]. Ciò è vero, in particolare, per scelte operate sotto pressione, senza sufficiente tempo a disposizione o in contesti a elevata complessità, tutte circostanze che fanno prevalere modalità decisionali basate sull'istinto e la rapidità su altre più lente, ma più razionali e ragionate. La teoria del *nudging* propone di investigare a fondo il contesto di riferimento in cui è presa una decisione, al fine di limitare gli effetti negativi di tali *bias* cognitivi. Per farlo, Thaler e Sunstein propongono di intervenire con procedimenti di *architettura della scelta* che rendano più probabile che un procedimento decisionale – anche eccessivamente rapido, approssimativo e istintivo – porti a un determinato risultato[298]. È un esempio di *nudge* il posizionamento di frutta e verdura nel punto più accessibile agli utenti di una mensa self-service, onde favorire un'alimentazione sana[299], o l'adozione un sistema di *opt-out* da un piano

[295]Cfr. ad esempio R. M. BOND, C. J. FARISS, J. J. JONES, A. D. I. KRAMER, C. MARLOW, J. E. SETTLE, J. H. FOWLER, *A 61-million-person experiment in social influence and political mobilization*, in *Nature*, 489, 2012, p. 295-298; Z. CORBYN, *Facebook experiment boosts US voter turnout*, in *Nature*, 2012, https://doi.org/10.1038/nature.2012.11401; D. SUSSER, B. ROESSLER, H. NISSENBAUM, *Technology, autonomy and manipulation*, in *Internet policy review*, 8, 2, 2019, DOI: 10.14763/2019.2.1410; J. KNOX, B. WILLIAMSON, S. BAYNE, *Machine behaviourism: future visions of "learnification" and "datafication" across humans and digital technologies*, in *Learning, Media & Technology*, 45, 1, 2020, p. 31-45, oltre ai già citati S. ZUBOFF, *The age of surveillance capitalism cit.*; K. YEUNG, *"Hypernudge": Big Data as a mode of regulation by design cit*.

[296]C. SUNSTEIN, R. THALER, *Nudge: improving decisions about health, wealth and happiness*, New Haven, 2008; cfr. ancheC. SUNSTEIN, *Nudging: a very short guide*, in *Journal of consumer policy*, 37, 2014, p. 583 ss.

[297] Si rinvia, in particolare, a D. KAHNEMAN, A. TWERSKY, *Prospect theory: an analysis of decision under risk*, in *Econometrica*, 47, 1979, p. 263-291; D. KAHNEMAN, A. TWERSKY, P. SLOVIC, *Judgment under uncertainity. Heuristics and biases*, Cambridge, 1982. Un altro contributo decisivo allo studio delle reali dinamiche di decisione in condizioni di incertezza è venuto dall'americano Vernon Smith, padre dell'economia sperimentale e vincitore, assieme a Kanheman, del premio Nobel per l'economia 2002. Cfr., in particolare, V. SMITH, *Rationality in economics: constructivist and ecological forms*, Leiden, 2007.

[298]Cfr. in particolare R. H. THALER, C. R. SUNSTEIN, *Nudge. The final edition*, New York, 2021, p. 91 ss.

[299] L'esempio è riportato dagli stessi R. H. THALER, C. R. SUNSTEIN, *Nudge. The final edition cit.*, p. 1 ss. ed è stato in seguito oggetto di diverse sperimentazioni, cfr. ad esempio M. VON ROOKHUIJZEN, E. DE VET, *Nudging healthy eating in Dutch sports canteens: a multi-method case study*, in *Public Health Nutrition*, 2020, doi:10.1017/S1368980020002013;

previdenziale aziendale al momento dell'assunzione, per garantire una vecchiaia serena alla maggior parte dei lavoratori[300]. La tesi di fondo è che l'opzione che si sceglie di favorire sarebbe quella che il soggetto preferirebbe se operasse una scelta perfettamente razionale, senza dubbio la migliore per lui o lei e la società nel suo complesso. Il *nudging* sarebbe preferibile a ogni forma di imposizione da parte dei poteri pubblici perché conserverebbe intatta la libertà di scelta, limitandosi a condizionarne i risultati. Thaler e Sunstein definiscono questo approccio *libertarian paternalism* e propongono una definizione di *nudge* in cui la libertà individuale ha un ruolo centrale: «*a nudge, as we will use the term, is any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates. Putting fruit at eye level counts as a nudge. Banning junk food does not*»[301].

È facile intuire come il *nudging* si presti ad applicazioni in ambito digitale. La stessa struttura di determinati servizi internet, in primo luogo le piattaforme social, è pensata per far trascorrere agli utenti più tempo possibile connessi – e dunque orientarne le decisioni[302]. L'elaborazione di dati personali attraverso l'intelligenza artificiale, inoltre, apre una nuova prospettiva: la possibilità di predisporre *nudge*personalizzati per un determinato profilo di utente, e, dunque, più efficaci[303]. Il comportamento dell'utente esposto al *nudge*, inoltre, rappresenta un *feedback* sul suo funzionamento, utilizzabile per il perfezionamento della profilazione. Tali *feedback*, peraltro, spesso sono chiesti esplicitamente dagli operatori di internet, che chiedono di valutare il gradimento di un determinato contenuto. A titolo di esempio, possono considerarsi tentativi di influenzare il comportamento dei soggetti interessati, riconducibili totalmente o parzialmente alla teoria del *nudging*: la selezione di un insieme di prodotti commerciali in un sito di*e-commerce*, ricavata dalle precedenti esperienze di acquisto; l'esposizione degli utenti di determinati servizi internet a contenuti e notizie orientati ideologicamente, con l'effetto di confermarne erafforzarne le convinzioni; la presentazione di un numero di contenuti audio e video "consigliati" nelle piattaforme *streaming*, scelti in base alle informazioni disponibili sui gusti dell'utente. La modifica

---

C. KAWA, P.M. IANIRO DAHM, J. F. H. NIJHUIS, W.H. GIJSELAERS, *Cafeteria online: nudges for healthier food choices in a university cafeteria – a randomized online experiment*.

[300] R. H. THALER, C. R. SUNSTEIN, *Nudge. The final edition cit.*, p. 18 ss.; Cfr. ancheR. L. CLARK, R. G. HAMMOND, M. SANDLER MORRILL, C. KHALAF, *Nudging retirement savings: a field experiment on supplemental plans*, Working Paper 23679 – National Bureau of Economic Research, Cambridge (USA), 2017;C.KRONCKE, *Nudging towards a stable retirement*, in *Politics and the Life Sciences*, 37, 1, 2018, p. 126-129.

[301] R. H. THALER, C. R. SUNSTEIN, *Nudge. The final edition cit.*, p. 11 ss.

[302]Cfr. ad esempio R. J. DEIBERT, *The road to digital unfreedom*, in *Journal of Democracy*, 30, 1, 2019, p. 25-39; V. R. BHARGAVA, M. VELASQUEZ, *Ethics of the attention economy: the problem of social media addiction*, in *Business Ethics Quarterly*, 31, 3, 2021, p. 321-359; J. FOX, *An unlikable truth: social media like buttons are designed to be addictive. They're impacting our ability to think rationally*, in *Index of censorship*, 47, 3, 2018, p. 11-13.

[303]Cfr. C. MELE, T. RUSSO SPENA, V. KAARTEMO, M. L. MARZULLO, *Smart nudging: how cognitive technologies enable choice architectures for value co-creation,* in *Journal for business research*, 129, 2021, p. 946-960; K. YEUNG, *"Hypernudge": Big Data as a mode of regulation by design cit.*

di scelte e preferenze generata da queste strategie può avere, ovviamente, un impatto anche sui comportamenti al di fuori dell'ambiente *online*. A tal proposito, deve segnalarsi che alcuni esperimenti sembrano dischiudere la possibilità di influenzare attraverso il *nudging* online singole condotte riguardanti la vita fuori dalla rete: è particolarmente famoso uno studio condotto dal *social network* Facebook, che, manipolando i contenuti cui erano esposte, avrebbe spinto a votare in un'elezione diverse migliaia di persone che altrimenti non l'avrebbero fatto[304].

Ciò che è di particolare interesse delle strategie di *architettura della scelta* basate sull'analisi dei dati appena viste è che esse non assumono, come obiettivo finale, il benessere dell'utente. L'interesse perseguito, infatti, è sempre quello del privato che fornisce il servizio, che analizza gusti e preferenze dell'utente e predispone il *nudge* per una finalità propria, spesso consistente nella massimizzazione dei profitti pubblicitari[305]. Si tratta di una circostanza assente nell'analisi di Thaler e Sunstein, che paiono dare per scontato, anche nelle edizioni più recenti del libro, che gli strumenti di *architettura della scelta* che teorizzano siano orientati a guidare, sotto il controllo dei poteri pubblici, i soggetti coinvolti verso i risultati migliori per loro stessi[306]. Quanto accade, invece, è di natura del tutto diversa, per la natura totalmente privata degli attori coinvolti e per gli scopi che perseguono.

2.3 Un possibile sviluppo ulteriore: i sistemi di credito sociale

Un ulteriore scenario reso possibile dalle tecniche più avanzate di elaborazione dei dati personali è rappresentato dai c.d. sistemi di credito sociale[307]. L'accumulazione di una mole sempre crescente di dati riguardanti gli individui nelle mani di poteri – in primo luogo pubblici, ma anche privati e informali –permette di catalogare quest'ultimi in funzione della loro aderenza a un determinato codice di condotta. La società digitale, infatti, rende agevole, in via teorica, accentrare una grande quantità di informazioni, raccolta dalle fonti più varie: social network, sistemi di pagamento

[304] Per dei commenti si vedano i già citati R. M. BOND, C. J. FARISS, J. J. JONES, A. D. I. KRAMER, C. MARLOW, J. E. SETTLE, J. H. FOWLER, *A 61-million-person experiment in social influence and political mobilization cit.*; Z. CORBYN, *Facebook experiment boosts US voter turnout cit.*

[305] Il riferimento, chiaramente, è in primo luogo alle riflessioni di S. ZUBOFF, *The age of surveillance capitalism cit.*

[306] Non si può non sottolineare come l'identificazione della scelta da preferire, in quanto nel miglior interesse del soggetto coinvolto, rimanga un'operazione complicata, che spesso implica giudizi di valore che sono, per loro natura, opinabili. Per questa e altre osservazioni sulla proposta di Sunstein e Thaler cfr., tra i molti, A. ALEMANNO, A. L. SIBONY (A CURA DI), *Nudge and the law: a european perspective*, Oxford, 2015; D. M. HAUSMAN, B. WELCH, *Debate: to nudge or not to nudge*, in *Journal of Political Philosophy*, 18, 1, p. 123-136.

[307] Cfr., in generale e da prospettive differenti, A. DEVEREAUX, L. PENG, *Give us a little social credit: to design or to discover personal ratings in the era of Big Data*, in *Journal of Institutional Economics*, 16, 2020, p. 369-387; Y. CHEN, A. S. CHEUNG, *The trasparent self under big data profiling: privacy and chinese legislation on the social credit system*, in *The journal of comparative law*, 12, 2, p. 356-378; R. CREEMERS, *China's social credit system: an evolving practice of control*, 2018, http://dx.doi.org/10.2139/ssrn.3175792 (8 febbraio 2022); S. HOFFMAN, *Managing the state: social credit, surveillance and CCP's plan for China*, in N. WRIGHT (A CURA DI), *AI, China, Russia and the global order: technological, political, global and creative*, Maxwell AFB, 2019, p. 48-55; R. BOTSMAN, *Who can you trust? How technology brought us together and why it might drive us apart*, New York, 2017; E. P. STRINGHAM (A CURA DI), *Private governance: creating order in economic and social life*, Oxford, 2015.

elettronico, cartelle cliniche elettroniche, procedimenti amministrativi digitalizzati ecc. L'analisi dei dati permette di trasformare tali informazioni in valutazioni sempre più precise e raffinate della corrispondenza dei cittadini al modello ideale di volta in volta prescelto. La sintesi può spingersi al punto di esprimere tale dato in forma numerica, assegnando all'individuo un punteggio in funzione della vicinanza alle caratteristiche considerate ottimali. Le tecnologie basate sull'apprendimento automatico permetterebbero di perfezionare e modificare costantemente il sistema e adeguare le sue valutazioni alle eventuali modifiche del comportamento dei soggetti coinvolti. Inoltre, i*feedback* con cui migliorare il sistema – ed è questo uno degli elementi più dirompenti della teorizzazione - potrebbero provenire direttamente dagli altri soggetti coinvolti, chiamati a esprimere una valutazione della condotta altrui. L'ottenere un punteggio alto o basso dovrebbe essere accompagnato da conseguenze diametralmente opposte: a un buon livello di fiducia e responsabilità dovrebbero corrispondere vantaggi in termini di accesso al credito, possibilità di rivestire cariche pubbliche e ruoli di potere, ecc. A punteggi bassi, invece, dovrebbero corrispondere svantaggi di varia natura o preclusioni apertamente sanzionatorie, come limitazioni alla possibilità di viaggiare o accedere a locali pubblici e l'esclusione da determinate professioni. Secondo i sostenitori della prospettiva, l'implementazione di un sistema di credito sociale come quello descritto rappresenterebbe un incentivo maggiore dei tradizionali sistemi di *enforcement* verso comportamenti virtuosi, responsabili e attenti ai bisogni della società[308]. L'efficacia, inoltre, sarebbe di gran lunga superiore a qualunque strategia di *nudging* digitale del tipo già analizzato, con le quali ha in comune l'idea di influenzare il comportamento umano attraverso l'analisi dei dati: il sistema, infatti, porterebbe a scegliere le condotte corrette grazie al timore di incorrere in pesanti e ben determinate conseguenze negative, invece di limitarsi a favorire determinate scelte, sfruttando comuni *bias* cognitivi.

Forme di valutazione numerica delle informazioni relative a un individuo da cui discendono precise conseguenze esistono da anni, spesso messe in atto da operatori privati: si pensi ai molti sistemi di *credit scoring*esistenti, il cui punteggio determina maggiori o minori possibilità di accedere al credito in funzione della propria storia personale[309]. Ciò nonostante, l'unico sistema di credito

[308] Per un esempio di questa visione cfr. A. ABBAS, J. KHALID, S. MUBARAK, H. JAVED, *Analyzing the reliability of human social scoring system (HSSS) & its determinants*, in *Journal of marketing and information systems*, 4, 1, p. 33-42 o, pur con alcuni accorgimenti giudicati necessari prima dell'adozione di tali sistemi nelle società liberali, J. MARGOLIS, *A Big Brother approach has qualities that would benefit society*, Financial Times, 31 ottobre 2017.

[309]Cfr. ad esempioH. A. ABDOU, J. POINTON, *Credit scoring, statistical techniques and evaluation criteria: a review of the literature,* in *Intelligent systems in accounting, finance and management*, 18, 2-3, p. 59-88; A. GHODSELAHI, A. AMIRMADHI, *Application of artificial intelligence techniques for credit risk evaluation*, in *International journal of modeling and optimization*, 1, 3, 2011, p. 243-249; L. XIAO-LIN, Y. ZHONG, *An overview of personal credit scoring: techniques and future work*, in *Internatonal Journal of Intelligence Science*, 2, 4A, 2012, DOI: 10.4236/ijis.2012.224024. Per un commento d'ambito giuridico, inoltre, v.L. AMMANNATI, G. L. GRECO, *Il credit scoring alla prova dell'intelligenza artificiale*, in U. RUFFOLO (A CURA DI), *XXXVI lezioni di diritto dell'intelligenza artificiale*, Torino, 2021, p. 373 ss.

sociale "completo", destinato cioè a raccogliere dati e ad avere conseguenze in una pluralità di ambiti dell'esistenza, è quello annunciato nel 2014 dalla Repubblica Popolare Cinese e finora messo in atto solo in modo frammentario. Nel 2014, infatti, il Consiglio di Stato cinese ha diffuso un documento intitolato, nella traduzione inglese, *Planning Outline for the Construction of a Social Credit System*, contenente un programma per la realizzazione di un sistema di credito sociale su scala nazionale entro il 2020[310]. Il piano, preceduto da alcune sperimentazioni a livello locale, indica quattro azioni fondamentali: sviluppare un quadro giuridico di supporto al sistema di credito sociale, creare sistemi omogenei di raccolta e verifica dei dati, rendere la fiducia un valore più rilevante nel mercato e articolare un corpo strutturato di benefici e sanzioni legate alle valutazioni del sistema[311]. La realizzazione concreta del progetto, in ogni caso, non ha rispettato i tempi previsti: ad oggi il sistema di credito sociale è implementato a macchia di leopardo sul territorio cinese, spesso coinvolgendo solo una parte dei cittadini, e si basa sull'inserimento in determinate *blacklist* sulla base di determinati comportamenti, e non sull'attribuzione di un punteggio numerico, raffinato costantemente, ai consociati in funzione delle loro azioni[312]. In ogni caso, già in questo stato iniziale possono derivare conseguenze negative da una pluralità di condotte, quali non pagare in tempo bollette e debiti, ascoltare musica ad alto volume, mangiare nella metropolitana o prenotare un hotel o ristorante e poi non presentarsi. Le sanzioni, a seconda della *blacklist* in cui si viene inseriti, implicano difficoltà nell'accesso al credito, nell'acquisto e nella locazione di immobili, limitazioni in viaggi e trasporti, l'esclusione da determinate professioni e cariche pubbliche e di eventuali figli dalle scuole più prestigiose[313].

[310] Una traduzione inglese del documento è disponibile in *Chinese copyright and media*, blog diretto da Rogier Creemes, docente di Modern Chinese Studies all'Università di Leiden, al link: https://bit.ly/3GTTrHU (16 gennaio 2022). Per alcuni commenti v. *China's corporate social credit system*, US Congressional research service – in focus, 17 gennaio 2020, https://bit.ly/3sEGcp6 (16 gennaio 2022). Per dei commenti si vedano, tra i molti, Y. CHEN, A. S. CHEUNG, *The trasparent self under big data profiling cit.*; R. CREEMERS, *China's social credit system cit.*; S. HOFFMAN, *Managing the state: social credit, surveillance and CCP's plan for China cit.*; S. ENGELMANN, M. CHEN, F. FISCHER, C. Y. KAO, J. GROSSKLAGS, *Clear Sanctions, Vague Rewards: How China's Social Credit System Currently Defines Good and Bad Behavior*, in *Proceedings of the Conference on Fairness, Accountability, and Transparency – ACM*, 2019, p. 69-78; K. HAO, *Is China's social credit system as Orwellian as it sounds?*, in *MIT Technology Review*, 26 febbraio 2018, https://bit.ly/3IWxdH0 (8 febbraio 2022); S. MISTREANU, *Life Inside China's Social Credit Laboratory*, in *Foreign Policy*, 3 aprile 2018, https://bit.ly/3MysgpZ(8 febbraio 2022).

[311] Cfr. *Planning Outline for the Construction of a Social Credit System*, nella citata traduzione inglese, https://bit.ly/3GTTrHU (8 febbraio 2022).

[312] Q. SUN, *China's social credit system was due by 2020 but is far from ready*, in *Algorithm Watch*, https://bit.ly/3pQavsC (8 febbraio 2022); L. MATSAKIS, *How the West got China's social credit system wrong*, in *Wired*, 29 luglio 2019, https://www.wired.com/story/china-social-credit-score-system/ (8 febbraio 2022); *China's Social Credit System in 2021: From fragmentation towards integration*, MERICS – Report, 3 marzo 2021, https://bit.ly/3KoGZSs (8 febbraio 2022).

[313]Cfr. *Planning Outline for the Construction of a Social Credit System cit.*; R. CREEMERS, *China's social credit system cit.*; S. MISTREANU, *China is implementing a massive plan to rank its citizens, and many of them want in*, in *Foreign Policy*, 3 aprile 2018, https://bit.ly/3pQcnBE (8 febbraio 2022); A. MA, *China's controversial social credit system isn't just about punishing people — herÈs what you can do to get rewards, from special discounts to better hotel rooms*, in *Business Insider*, 3 febbraio 2019, https://bit.ly/3MKITz1 (8 febbraio 2022)*;* A. F. ELLIOT,*China is banning people with*

Infine, per quanto quello cinese rimanga, allo stato dell'arte, l'unico sistema di credito sociale effettivamente in funzione, seppur parzialmente rispetto alle intenzioni dichiarate, deve segnalarsi che anche Venezuela e Russia, rispettivamente nel 2017 e nel 2018, hanno esplicitato l'intenzione di sviluppare forme di "cittadinanza digitale" accostate da più parti al progetto cinese[314]. Analogie sempre più profonde coi sistemi di credito sociale, inoltre, sono state identificate coi già menzionati strumenti di *credit scoring*, utilizzati da privati, in versionisempre più versatili e sofisticate, in vari paesi del mondo, comprese diverse democrazie consolidate[315].

3. Le sfide poste dall'intelligenza artificiale alle varie dimensioni dell'identità personale: tutela della riservatezza, controllo sui dati personali, diritto all'oblio e questioni aperte

3.1 L'evoluzione del diritto all'identità personale nell'era digitale: dalla riservatezza al controllo sui dati personali

È stato già evidenziato che giurisprudenza e dottrina, in Italia e altrove, hanno inteso la protezione dell'identità principalmente come diritto a una libera espressione e corretta rappresentazione pubblica della personalità. Invece, la libertà di definire tale personalità senza indebite interferenze, antecedente necessario di quest'ultima,è stata oggetto di minor elaborazione teorica, risultando nei fatti ampiamente garantita dai diritti e dalle libertà riconosciuti dalla Carta fondamentale e dalle tradizioni costituzionali più vicine alla nostra. Vi sono state, in ogni caso, alcune vistose eccezioni, in cui si è resa necessaria l'elaborazione di strumenti giuridici innovativi per proteggere l'intimità e la libertà dell'individuo di definirsi e determinarsi da sé. Il mutamento dei costumi e l'evoluzione tecnologica, infatti, hanno generato situazioni di fronte alle quali l'impostazione tradizionale pareva offrire una tutela incerta. In prima battuta, è stato il caso della riservatezza, intesa come diritto a tenere nascosti determinati aspetti della propria esistenza. I giudici italiani, al pari dei colleghi di altri paesi dell'Europa continentale, hanno cominciato a dar tutela a tale nuova posizione giuridica a partire dagli anni '50 del XX secolo, con più di cinquant'anni di ritardo rispetto agli ordinamenti di common law e, in particolare, agli Stati Uniti[316]. Le prime pronunce, da ambo i lati dell'Atlantico,

---

bad 'social credit' from using planes and trains, in *The Telegraph*, 19 maggio 2018, https://bit.ly/3sRRWWE (8 febbraio 2022).

[314] A. BERWICK, *Venezuela is rolling out a new ID card manufactured in China that can track, reward and punish citizens*, in *Business Insider*, 18 novembre 2018, https://bit.ly/3pNC5H0 (8 febbraio 2022); *80% of Russians Will Have State-Gathered 'Digital Profiles' by 2025, Official Says*, in *The Moscow Times*, 28 settembre 2018, https://bit.ly/3hTaXBX (8 febbraio 2022).

[315] Cfr. ad esempio Z. WILLIAMS, *Algorithms are taking over – and woe betide anyone they class as a 'deadbeat'*, The Guardian, 12 luglio 2018; *Warning: Germany edges toward Chinese-style rating of citizens*, Handelsblatt Global Edition, 17 February 2018.

[316] Tradizionalmente, la prima compiuta elaborazione teorica del diritto si fa risalire a un notissimo articolo dei giuristi americani Samuel Warren e Luis Brandeis, pubblicato nel 1890 sull'*Harward Law Review*, cfr. S. WARREN, L. BRANDEIS, *The right to privacy*, in *Harvard Law Review*, 4, 5, 1890, p. 193-220; il primo riconoscimento legislativo, sullo scenario statunitense, risale al 1903, con l'emanazione, da parte dello stato di New York, di due *statute* che

riguardavano le vicende di personaggi noti, la cui vita privata era stata fotografata, commentata e diffusa al pubblico a loro insaputa dalla stampa scandalistica[317]. In gioco, dunque, non vi era la verità della propria rappresentazione pubblica, ma la possibilità di escludere da essa taluni elementi, perché considerati intimi e comunque compromettenti (il diritto, secondo una nota espressione inglese, a*belet alone*[318]). La riservatezza ha, chiaramente, un valore fondamentale dal punto di vista del libero sviluppo della propria personalità, che può dirsi pieno solo quando esista la garanzia di poter tenere nascoste determinate informazioni su sé stessi, esattamente come le mura domestiche custodiscono la libertà di ciò che avviene al loro interno[319].

Come già detto, l'avvento degli elaboratori elettronici ha complicato di molto le cose. La semplice riservatezza, quale diritto a non vedere esposti fatti puntuali della propria esistenza, risultava del tutto insufficiente di fronte alle questioni sollevate dalla possibilità di memorizzare e analizzare dati personali in modo automatizzato. Si poneva il nuovo problema dell'uso che operatori del mercato e autorità di pubblica sicurezza avrebbero potuto fare della mole di informazioni disponibili sugli

---

riconoscevano il diritto di agire a tutela dell'interesse alla privacy; il primo riconoscimento pieno da parte della giurisprudenza del diritto alla privacy è avvenuto invece nel 1905 da parte della Corte Suprema della Georgia, nel caso *Pavesich vs. New England Life Insurance Company*; la tutela della privacy poteva dirsi più che consolidata nel sistema americano già nel 1939, quando l'American Law Institute vi dedicava una parte del *Restatement of Torts*. Ancora precedente è invece il riconoscimento di una forma di tutela alla riservatezza nell'ordinamento inglese, risalente almeno al caso *Prince Albert v. Strange* (1849) 1 Mac and G 25, 1H e TW1, *Court of Chancery*, in cui è sanzionato l'illecito di *sordidspying in the privacy of domestic life*, cfr. A. CERRI, *Riservatezza (diritto alla) II – diritto comparato e straniero*, in *Enciclopedia giuridica*, XXVII, 1995. Nell'ordinamento italiano servirà invece attendere il secondo dopoguerra, prima col *caso Petacci*, cfr. Cass. civ., 20 aprile 1963, n. 990, in *Foro it.*, 1963, I, p. 877, con nota di A. DE CUPIS,*Riconoscimento sostanziale, ma non verbale, del diritto alla riservatezza*, poi, senza più ambiguità, col *caso Soraya*, cfr. Cass. civ., 27 maggio 1975, n. 2129, in *Foro it.*, 1976, I, p. 2895, con nota di M. MONTELEONE, dopo l'iniziale diniego dell'esistenza stessa di un diritto risarcibile alla riservatezza nel *caso Caruso*, cfr. Cass. 22 dicembre 1956, n. 4487, in *Foro it.*, 1957, I, p. 4, con nota di A. DE CUPIS, *Sconfitta in Cassazione del diritto alla riservatezza*, p. 232 ss., e in *Giur. it.* 1957, I, p. 366, con nota di G. PUGLIESE, *Una messa a punto della Cassazione sul preteso diritto alla riservatezza*.

[317] È il caso dello stesso Samuel Warren, noto avvocato bostoniano che, dopo aver sposato la figlia di un facoltoso uomo politico, cominciò a condurre una vita mondana e dispendiosa, attirando presto l'interesse della nascente stampa scandalistica. Fu proprio la diffusione a mezzo stampa, per l'ennesima volta, di fatti della vita sua vita privata a spingerlo a pubblicare il noto articolo del 1890 insieme all'amico Luis Brandeis, futuro giudice della Corte Suprema. La vicenda all'origine del caso *Prince Albert v. Strange* riguardava invece dei ritratti ritraenti la famiglia reale inglese, che un dipendente della casa reale aveva copiato abusivamente e progettava di pubblicare. In Italia, il *caso Petacci* atteneva alla violazione della riservatezza della famiglia Petacci, a seguito della pubblicazione di un volume narrante dettagli della vita privata dell'amante di Benito Mussolini, Claretta Petacci, e dei suoi congiunti; il *caso Soraya* riguardava la lesione della riservatezza di Soraya Esfandiary, ex moglie dell'ultimo Mohammad Reza Pahlavi, ritratta dalla stampa intima in atteggiamenti intimi con un altro uomo all'interno delle mura domestiche.

[318] L'espressione, coniata da Justice Thomas M. Cooley nel suo *Law of torts* del 1880, è utilizzata dagli stessi S. WARREN, L. BRANDEIS, *The right to privacy cit.*, p. 195: «Recent inventions and business methods call attention to the next step wich must be taken for the protection of the person, and for securing to the individual what Judge Cooley calls the right *to be let alone* [corsivo aggiunto]», cfr. anche T. M. COOLEY, *A Treatise on the Law of Torts or the Wrongs Which Arise Independently of Contract*, Chicago, 1880, par. 29.

[319] Si rimanda, tra i molti e senza animo di completezza, agli studi di S. RODOTÀ, *Tecnologie e diritti*, Bologna, 1995; *Privacy, libertà, dignità*, Discorso conclusivo della XVI Conferenza internazionale sulla protezione dei dati, Wroclaw (PL), 14, 15, 16 settembre 2004; *Intervista su privacy e libertà* (A CURA DI P. CONTI), Bari, 2005; T. A. AULETTA, *Riservatezza e tutela della personalità*, Milano, 1978; A. BALDASSARE, *Privacy e costituzione*, Roma, 1974.

individui, anche senza renderle pubbliche[320]. Ciò ha portato al riconoscimento, in varie forme, del già menzionato diritto al controllo sui dati personali, consistente nella possibilità di conoscere chi sia in possesso di determinate informazioni, archiviate elettronicamente, e di influire su di esse[321]. Nella varietà di soluzioni reperibili sul piano globale, l'approccio europeo al controllo sui dati personali appare il più garantista, escludendo, dal punto di vista concettuale, che i dati personali siano considerabili alla stregua di un mero bene commerciale (tanto che vi è inapplicabile l'istituto della proprietà[322]) e prevedendo che il loro trattamento avvenga per finalità determinate e sulla base di presupposti di liceità esplicitamente individuati dalla legge[323]. Nel mercato dei servizi internet, assume un ruolo centrale il consenso dell'individuo interessato, sia come autonoma base giuridica del trattamento, sia dal punto di vista di un eventuale accordo contrattuale stipulato on-line, che rende lecito, sempre ai sensi delle norme in materia, il trattamento dei dati necessario a darvi esecuzione[324].L'impostazione che emerge dall'attuale quadro normativo europeo, quindi, legittima la raccolta di quantità sempre crescenti di dati personali che oggi avviene online in larga parte sulla

[320]Per queste riflessioni cfr. in primo luogo S. RODOTÀ, *Elaboratori elettronici e controllo sociale cit.; Tecnologie e diritti cit.* Inoltre, si vedano *ex multis* G. PASCUZZI, *Il diritto dell'era digitale*, Bologna, 2020, p. 47 ss; G. D. FINOCCHIARO, *La protezione dei dati personali e la tutela dell'identità*, in *Diritto di Internet*, Bologna, 2020, p. 151-183.

[321] Per una panoramica delle tempistiche d'introduzione e dell'evoluzione successiva della normativa in materia di dati personali nei vari ordinamenti cfr. la giá citata F. MOLNAR-GABOR, *Data Protection cit.*

[322] Per le criticità di questa impostazione si rimanda al lavoro di N. PURTOVA, *Property in personal data: a European perspective on the instrumentalist theory of propertization*, in *Law and technology – selected essays*, Pistoia, 2009, p. 225-243; *Property rights in personal data: learning from the American discourse*, in *Computer law & security review*, 25, 6, 2009, p. 507-521; *The illusion of personal data as no ones property*, in *Law, innovation & technology*, 7, 1, 2015, p. 83-111, doi: 10.1080/17579961.2015.1052646.

[323]Cfr. R. WALTERS, L. TRAKMAN, B. ZELLER, *Data protection law. A comparative analysis of Asia-Pacific and European approaches*, Singapore, 2019, p. 79-81; v. anche i contributi raccolti in D. DÖRR, R.L. WEAVER (A CURA DI), *Perspectives on privacy:increasing regulation in the USA, Canada, Australia and European countries*, Berlino-Boston, 2014. Il GDPR elenca le basi giuridiche legittimanti il trattamento dei dati personali all'art. 6 e, per quanto riguarda le categorie particolari di dati personali (largamente accostabili ai c.d. *dati sensibili* previsti dal precedente regime) il cui trattamento è previsto solo in via d'eccezione, all'art. 9.

[324]Cfr. in generale E. KOSTA, *Consent in European data protection law*, Leiden, 2013. L'art. 6 par. 1 del GDPR, infatti, indica come prima base giuridica, alla lett. a): «l'interessato ha espresso il consenso al trattamento dei propri dati personali per una o più specifiche finalità»; cui segue, alla lett. b): «il trattamento è necessario all'esecuzione di un contratto di cui l'interessato è parte o all'esecuzione di misure precontrattuali adottate su richiesta dello stesso». Un terzo presupposto di liceità spesso indicato dalle piattaforme per il trattamento dei dati dei loro utenti è il c.d. legittimo interesse, previsto alla lett. f) del menzionato art. 6 par. 1: «il trattamento è necessario per il perseguimento del legittimo interesse del titolare del trattamento o di terzi, a condizione che non prevalgano gli interessi o i diritti e le libertà fondamentali dell'interessato che richiedono la protezione dei dati personali, in particolare se l'interessato è un minore». Anche in quest'ultimo caso, però, l'adesione consensuale dell'utente ai servizi offerti dal *provider* è l'antecedente logico necessario che rende possibile il trattamento e, dunque, l'elemento volontaristico appare, dal punto di vista sostanziale, ancora determinante. A titolo d'esempio, *Meta* dichiara nella privacy policy dei servizi Facebook, Instagram e Messenger come principale base giuridica l'esecuzione di obblighi contrattuali, cui seguono, in ordine d'importanza, il consenso dell'interessato (ad es. quando siano coinvolte categorie particolari di dati personali, per le quali l'art. 9 GDPR non prevede l'esecuzione di obblighi contrattuali come condizione di liceità) e il legittimo interesse (ad es. quando il trattamento riguardi dati di minori d'età, la cui limitata capacità di contrarre in quasi tutti gli ordinamenti potrebbe rendere inutilizzabile la basse giuridica dell'esecuzione di obblighi contrattuali), cfr: https://www.facebook.com/about/privacy/legal_bases (28 febbraio 2022). Per un approfondimento del quadro giuridico applicabile al trattamento dati nel contesto della fornitura di servizi internet in Europa si rimanda alle Linee Guida dell'EDPD n. 2 del 9 aprile 2019, *On the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects*, https://bit.ly/3tN6gjG (28 febbraio 2022).

base dellavolontà degli individui interessati, della cui genuinità, come vedremo, si può dubitare. Lo scenario non cambia negli Stati Uniti, dove hanno sede gran parte delle principali piattaforme, nonostante la notevole diversità della normativa di riferimento. Negli USA, infatti, predomina un approccio mercantilista e proprietario alla circolazione dei dati personali, e non esiste un testo normativo unitario in materia[325]. La tutela dei diritti degli individui cui i dati si riferiscono è affidata a discipline settoriali, codici di autoregolamentazione, e, in primo luogo, allo strumento contrattuale, accedendo al quale i cittadini accettano le modalità di trattamento di volta in volta adottate[326]. È evidente, anche in questo caso, il ruolo centrale attribuito alla volontà.

3.2 Il diritto all'evoluzione della propria identità e le nuove tecnologie: le vicende del diritto all'oblio

Le elaborazioni in materia di riservatezza e di dati personali miravano a proteggere la libera formazione dell'identità, limitando il condizionamento causato dall'incontrollata esposizione pubblica di fatti privati o da forme di schedatura elettronica al di fuori di espliciti agganci normativi. Alcune note vicende giurisprudenziali degli ultimi decenni, invece, mettono in luce come l'attuale struttura dei servizi internet possa rappresentare un ostacolo anche per la possibilità di modificare e far evolvere la propria identità. Negli ultimi tre decenni, infatti, cittadini di diversi paesi si sono rivolti ai rispettivi tribunali nazionali, chiedendo che determinate vicende del loro passato, che avevano attirato l'attenzione della stampa, venissero rese meno agilmente reperibili sul web. L'attenzione si è presto spostata dai periodici nei cui archivi online tali informazioni erano reperibili ai motori di ricerca che permettevano al pubblico di accedervi con facilità. Nella quasi totalità dei casi, i loro sistemi di indicizzazione dei contenuti – che generalmente impiegano tecnologie basate sull'intelligenza artificiale – presentavano le notizie in questione tra i primi risultati anche in seguito aricerche estremamente generiche, come il semplice nome e cognome del soggetto coinvolto. Com'è intuibile, la delicatezza degli interessi in esame, che impone di bilanciare, da un lato, libertà d'espressione e diritto di cronaca e, dall'altro, diritti della personalità dell'interessato, ha portato a una grande varietà di soluzioni, anche tra corti dello stesso paese[327].

[325] In materia, cfr. in particolare D. SOLOVE, P. M. SCHWARTZ, *Information privacy law (7th ed.)*, New York, 2021; W. MCGEREVAN, *Privacy and data protection law*, St. Paul, 2016.

[326] Tra i principali testi normativi in materia di *information privacy* devononominarsi, in particolare, a livello federale l'*Health Insurance Portability and Accountability Act* del 1996, il*Children's Online Privacy Protection Act* del 1998 e il *Fair and Accurate Credit Transactions Act* del 2003. Per quanto riguarda la legislazione dei singoli stati, spiccano le iniziative della California, in netta controtendenza rispetto all'impostazione generale statunitense, e in particolare il *California Online Privacy Protection Act* del 2003 e il *California Consumer Privacy Act* del 2018. La California, inoltre, è l'unico stato americano a riconoscere esplicita protezione al *right to privacy* già nel primo articolo della propria costituzione.

[327]In Italia, ad esempio, le prime sentenze a riconoscere il diritto all'oblio risalgono almeno agli anni '90. È del 1995, in particolare, un noto pronunciamento del Tribunale di Roma riguardante la ripubblicazione, nel 1990, della prima pagina del 7 dicembre 1961 da parte di un noto quotidiano, nell'ambito di un gioco a premi rivolto ai lettori. La testata

Un definitivo cambio di passo, sullo scenario europeo, è stato causato dalla nota sentenza del 2014 della Corte di Giustizia sul caso *Google Spain*[328]. Adita in via pregiudiziale dall'*Audiencia Nacional Española* relativamente al caso di un cittadino che lamentava la troppa facilità con cui, attraverso il motore di ricerca, era reperibile un articolo di giornale di diversi anni prima relativo a un procedimento esecutivo da egli subito, la Corte ha riconosciuto il diritto a non vedere reperibili online con eccessiva facilità notizie non più attuali, qualora ciò non appaia giustificato da interessi

---

riportava foto e nome di un soggetto reo confesso di omicidio, che trent'anni dopo aveva completamente scontato la pena e si era positivamente reinserito nella società. Il Tribunale ha statuito che mancasse, nel caso specifico, qualunque interesse attuale alla ridiffusione della notizia e ha condannato il giornale al risarcimento del danno (giungendo, peraltro, a riconoscere contestualmente gli estremi del reato di diffamazione a mezzo stampa), cfr. Tribunale Roma, 15 maggio 1995, in *Diritto di famiglia e delle persone*, 27, 1, 1998, con nota di G. CASSANO, p. 76 ss. L'anno dopo, invece, lo stesso Tribunale di Roma rifiutò di riconoscere tutela alle pretese dei familiari di una vittima di un noto omicidio degli anni '70, del quale la RAI intendeva proporre la ricostruzione in una puntata del programma *I grandi processi*, ritenendo che la ridivulgazione fosse giustificata da un chiaro interesse sociale, cfr. Tribunale di Roma, 27 novembre 1996, in *Giur. cost.* 1997, con nota di A. MASARACCHIA, p. 3018 ss. La stessa sorte ebbe, nel 2001, il ricorso d'urgenza intentato di fronte al medesimo Tribunale di Roma da una persona legata sentimentalmente, all'epoca dei fatti, a un membro della c.d. banda della Uno Bianca, volto a inibire la messa in onda di uno sceneggiato televisivo da parte della rete privata Canale 5, cfr. Tribunale Roma, 1 febbraio 2001, in *Diritto dell'informazione e dell'informatica*, 2001, p. 206. Il tema ha interessato più volte anche il Garante della privacy, che, già nel 2004, aveva individuato in un suo provvedimento una soluzione molto simile a quella poi adottata dalla Corte di Giustizia UE, prevedendo che il nome di un soggetto incorso, in passato, in una sanzione erogata da un ente pubblico potesse rimanere esposto al pubblico in un'apposita pagina del sito internet dell'ente in questione, ma dovesse essere reso irraggiungibile attraverso i più comuni motori di ricerca, v. GPDP, Dec. 10 novembre 2004, https://bit.ly/3iYEbjn (1 marzo 2022). La Cassazione, nel 2012, aveva poi chiarito che condizione di liceità della ridiffusione di notizie del passato è la sussistenza di un interesse attuale alla loro conoscibilità, e che anche di fronte alla permanenza di tale interesse, i media che rendano reperibili vicende risalenti nel tempo devono parimenti fornire informazioni su loro eventuali aggiornamenti (il caso riguardava la notifica dell'arresto di un noto personaggio politico, poi assolto anni dopo), cfr. Cass. civ. sez. III, 5 aprile 2012, n. 5525, con nota di A. MANTELEROin*La Nuova Giurisprudenza Civile Commentata*, 10, 2012, p. 843 ss. L'idoneità del diritto all'oblio a limitare il diritto di cronaca era stata poi confermata dalla Suprema Corte l'anno successivo. Per quanto riguarda la Germania, hanno suscitato particolare scalpore le richieste a vari media, compresa Wikipedia, dei responsabili dell'omicidio del noto attore Walter Sedlmair, volte a vedere i loro nomi eliminati dagli archivi digitali riportanti notizie sul caso. Le corti tedesche hanno sposato soluzioni differenti: i giudici di primo grado di Amburgo, nel 2008, hanno riconosciuto tutela inibitoria ai due ricorrenti, ingiungendo al periodico *Spiegel*di rimuovere i loro nominativi, v. LG Hamburg, Urteilvom 18.01.2008 - 324 O 507/07, https://oj.is/371835 (1 marzo 2022). La decisione, confermata in appello, è stata ribaltata in terzo grado dalla Corte federale di Karlsruhe, che, nel 2009, ha rifiutato la richiesta di rimozione dei due nomi e cognomi anche in un caso analogo, attiente alla banca dati online dell'emittente *Deutschlandradio* (soluzione poi confermata nel 2018, anche dalla Corte EDU), cfr. BGH, Urteilvom 09.02.2010 - VI ZR 244/08; BGH, Urteilvom 15.12.2009 - VI ZR 227/08; ECHR 554 M.L. e W.W. v. Germany 60798/10 (28 giugno 2018). Una strada ancora diversa era stata tentata dalla Francia, che aveva cercato di disciplinare il bilanciamento del diritto all'oblio con gli altri interessi coinvolti attraverso la strada della co-regolazione, stipulando con un'ampia varietà di operatori del web, nel 2010, la *Chartedudroit à l'oublidansles site collaboratifs et lesmoteurs de recherche* (tra i cui firmatari non risultavano, però, player chiave come Google o Facebook), https://bit.ly/3iXjJzb (1 marzo 2022). Diverso, come si dirà (cfr. n. 331) è l'atteggiamento della dottrina statunitense: i giudici americani, in particolare, si sono dimostrati particolarmente restrittivi in materia di riconoscimento del diritto all'oblio, dopo alcune aperture risalenti nel tempo.

[328] Corte di Giustizia dell'Unione Europea (Grande Sezione), 13 maggio 2014, *Google Spain e Google Inc. c. Agencia Española de Protección de Datos e Mario Costeja González*, C-131/12. Tra i numerosissimi commenti alla pronuncia si segnalano, senz'animo di completezza, O. LYNSKEY, *Control over Personal Data in a Digital Age: Google Spain v AEPD and Mario Costeja Gonzalez*, in *The modern law review*, 78, 3, 2015, p. 522-534; S. KULK, F. ZUIDERVEENBORGESIUS, *Case Notes: Google Spain vs. González: did the Court forget about freedom of expression?*, in *European Journal of Risk Regulation*, 3, 2014, p. 389-398; G. SARTOR, *Search engines as controllers. Inconvenientimplications of a questionableclassification*, in *Maastricht Journal of European Comparative Law*, 21, 3, 2014, p. 564-575; E. STRADELLA, *Cancellazione e oblio: come la rimozione del passato, in bilico tra tutela dell'identità personale e protezione dei dati, si impone anche nella Rete, quali anticorpi si possono sviluppare, e, infine, cui prodest?*, in *Rivista AIC*, 4, 2016, e i contributi raccolti in G. RESTA, V. ZENO-ZENCOVICH, *Il diritto all'oblio su internet dopo la sentenza Google Spain*, Roma, 2015.

di pari valore. La sentenza, inoltre, ha stabilito che l'attività di indicizzazione svolta dai motori di ricerca rientra nella definizione di trattamento dei dati personali elaborata dal diritto europeo e ha imposto a tali operatori di predisporre dei meccanismi con cui i privati possano chiedere la deindicizzazione di determinati contenuti online che li riguardino. Tale diritto a non essere perseguitati dal passato e a poter evolvere la propria personalità, senza veder inficiato ogni tentativo in tal senso da una superficiale ricerca online da parte di chiunque, ha assunto, nel lessico giuridico, l'evocativo nome di diritto all'oblio o, nella versione inglese, *right to be forgotten*. È stato, in seguito, esplicitamente riconosciuto dall'art. 17 del GDPR, in fase di discussione presso le istituzioni europee mentre veniva emanata la sentenza *Google Spain*[329]. Deve segnalarsi che, allo stato dell'arte, la soluzione elaborata dalla Corte di Giustizia è un *unicum* sullo scenario mondiale: corti di diversi ordinamenti extraeuropei hanno riconosciuto, in vicende puntuali, la risarcibilità del danno dovuto all'eccessiva e ingiustificata ridiffusione di notizie non più attuali, ma nessun altro ordinamento ha elaborato un meccanismo simile al procedimento di deindicizzazione appena descritto[330]. Le corti degli Stati Uniti, in particolare, si sono dimostrate estremamente restrittive in materia di diritto all'oblio, considerato dalla dottrina prevalente incompatibile con il Primo Emendamento[331]. La stessa Corte di Giustizia, d'altronde, adita in via pregiudiziale nel 2019 dal Conseil d'État francese, in merito a una controversia che vedeva protagonista, ancora una volta, il motore di ricerca *Google*, ha chiarito che gli obblighi in materia di deindicizzazione si applicano solamente sul territorio dell'Unione Europea[332].

[329] L'inizio dell'art. 7 par. 1 GDPR, infatti, recita: «L'interessato ha il diritto di ottenere dal titolare del trattamento la cancellazione dei dati personali che lo riguardano senza ingiustificato ritardo e il titolare del trattamento ha l'obbligo di cancellare senza ingiustificato ritardo i dati personali». A questa enunciazione di principio segue l'elenco delle condizioni che legittimano l'esercizio del diritto. Cfr. anche le Linee Guida n. 5/2019 dell'EDPD, *On the criteria of the right to be forgotten in these archengines cases unde rthe GDPR*, https://bit.ly/3IYbY6D (1marzo 2022).

[330] Hanno riconosciuto tutela al diritto all'oblio, sul piano risarcitorio e ordinando la rimozione di determinati contenuti dai siti web coinvolti, ad esempio, le corti di Argentina o India, senza però, caricare di oneri per il futuro i motori di ricerca, v. E. L. CARTER, *Argentina's right to be forgotten*, in *Emory International Law Review*, 27, 2013, p. 23-41; *Indiansfight for the "right to be forgotten online"*, AlJazeera, 16 marzo 2022 https://bit.ly/3iVLHLP (20 marzo 2022); *Derecho al olvido: historico fallo contra Google en elpais*, Cadena3, 11 agosto 2020, https://bit.ly/3uOxKol (20 marzo 2022); *Para la Corte Suprema, losbuscadores no son responsables del contenidoquelistan*, La Nación, 14 gennaio 2015, https://bit.ly/3NFzM2T (20 marzo 2022).

[331] Infatti, nonostante alcune aperture, anche piuttosto risalenti nel tempo, da parte dei giudici americani – cfr. in particolare *Melvin v. Reid*, 112 Cal.App. 285, 297 p. 91 (1931); *Briscoe v. Reader's Digest* 483 P.2d34 (Cal. 1971) - la Corte Suprema ha negato in diverse occasioni l'esistenza di un diritto all'oblio azionabile di fronte ai media, cfr. ad es.U.S. Supreme Court, *CoxBroadcasting v. Cohn*, 420 U.S. 469 (1975); U.S. Supreme Court, *Florida Star v. B.J.F.*, 491 U.S. 524 (1989); U.S. Supreme Court *Bartnicki v. Vopper*, 532 U.S. 514 (2001); In letteratura v., da punti di vista differenti, J. ROSEN, *The right to be forgotten*, in *Stanford Law Review Online*, 88, 64, 2012, http://www.stanfordlawreview.org/online/privacy-paradox/right-to-be-forgotten (4 marzo 2021) e *The unvanted gaze*. *The destruction of privacy in America*, New York, 2000; A. GAJDA, *Privacy, press and the right to be forgotten in the United States*, in *Washington Law Review*, 93, 201, 2018;R.ANTANI, *The resistance of memory: could the European Union's right to be forgotten exists in the United States?*, in *Berkeley Technology Law Journal*, 30, 385, 2015, p. 1173-1210 ss.

[332] Corte di Giustizia dell'Unione Europea (Grande Sezione), 24 settembre 2019, *Google LLC. c. Commissionnationale de l'informatique et deslibertés*, C-507/2017. In letteraturasirimanda, *ex multis*, a M. ZALNIERIUTE, *Google LLC v. Commission nationale de l'informatique et des libertés (CNIL)*, in *American Journal of International Law*, 114, 2, 2020, p. 261-267, P. T. J. WOLTERS, *The territorial effect of the right to be forgotten after Google v CNIL*, in *International*

3.3 Tutele giuridiche non al passo coi tempi: le questioni aperte dall'intelligenza artificiale per la protezione del diritto all'identità personale

Per quanto significative, le elaborazioni legislative, dottrinali e giurisprudenziali in materia di riservatezza, protezione dei dati personali e diritto all'oblio appena viste cercavano di dare risposta a minacce circoscritte per l'identità individuale, rese possibili dall'innovazione tecnologica e verso le quali il quadro giuridico di riferimento appariva inadeguato. In questa sede, preme ora evidenziare che la diffusione dell'intelligenza artificiale rende tali strumenti giuridici del tutto insufficienti per tutelare l'identità della persona. Infatti, l'analisi dei dati con strumenti di *machine learning*, la diffusione capillare dell'*IoT*, l'integrazione di componenti, hardware e software, basati sull'intelligenza artificiale nelle tecnologie d'uso comune hanno aumentato e diversificato i rischi per il bene giuridico dell'identità e i diritti ad esso collegati. La memorizzazione e l'elaborazione automatizzata di dati personali conducono spesso a risultati non prevedibili per l'essere umano. La precisione di un'attività di profilazione, ad esempio, può superare ogni aspettativa, e da pochi dati noti, aggregati con altri riferiti ad altri soggetti, è possibile ricavare informazioni del tutto inaspettate[333]. Di fronte a tali situazioni, il modello del consenso, che, come già visto, sovente funge, in via diretta o indiretta, da base di legittimazione per i trattamenti di dati personali online, mostra tutta la sua inadeguatezza, chiedendo all'utente, nella pratica, di accettare o no conseguenze che non può essere in grado di prevedere[334]. Inoltre, la frequenza con cui all'utente di servizi internet è richiesto di esprimere il proprio consenso, o comunque di prendere visione delle modalità

Journal of Law and Information Technology, 29, 1, p. 57-75; J. GLOBOCNIK, *The Right to Be Forgotten is Taking Shape: CJEU Judgments in GC and Others (C-136/17) and Google v CNIL (C-507/17)*, in *GRUR International – Journal of European and International IP Law*, 69, 4, p. 380-388; S. WRIGLEY, A. KLINEFELTER, *Google LLC vs. CNIL: the location-based limits of the EU right to erasure and lessons for U.S. privacy law*, in *North Carolina Journal of Law & Technology*, 22, 4, 2021, p. 681-735; F. BALDUCCI ROMANO, *La Corte di giustizia "resetta" il diritto all'oblio*, in *Federalismi.it*, 3, 2020, p. 31-46; R. PARDOLESI, *(Protezionedeidatipersonali) Nota a sent. CGUE grande sez. 24 settembre 2019 (causa C-507/17 Google France vs CNIL); sent. CGUE grande sez. 24 settembre 2019 (causa C-136/17)*, in *Foro italiano*, 2019, 12, 4, p. 594-597; O. POLLICINO, *L' "autunno caldo" della Corte di giustizia in tema di tutela dei diritti fondamentali in rete e le sfide del costituzionalismo alle prese con i nuovi poteri privati in ambito digitale*, in *Federalismi.it*, 19, 2019, p. 2-15.

[333] Cfr. ad esempio K. MANHEIM, L. KAPLAN, *Artificial intelligence: risks to privacy and democracy*, in *Yale Journal of Law & Technology*, 21, 2019, p. 106-188; C. TUCKER, *Privacy, algorithms and artificial intelligence*, in *The economics of artificial intelligence: an agenda*, Chicago, 2019, p. 423-437; T. TIMAN, Z. MANN, *Data protection in the era of artificial intelligence: trends, existing solutions and recommendations for privacy-preserving technologies*, in E. CURRY ET AL. (A CURA DI), *The elements of big data value. Foundations of the research and innovation ecosystem*, Cham, 2021, p. 153-177; RAFFIOTTA E., *Artificial intelligence, identification tools and identity protection*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2022, p. 165-179.

[334] Cfr., tra i molti, gli studi di D. J. SOLOVE, *Privacy self-management and the consent dilemma*, in*Harvard Law Review*, 126,2013,p.1880 – 1903; N. BROCKDORFF, S. APPLEBY-ARNOLD, *What consumers think*, EU CONSENT Project, Workpackages 7-8, 2013; R. BÖHME, S. KÖPSELL,*Trained to accept? A field experiment on consent dialogs*, in*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010, p. 2403–2406; B. W. SCHERMER, B. CUSTERS, S. VON DER HOF, *The crisis of consent: how stronger legal protection may lead to weaker consent in data protection*, in *Ethics and information technology*, 16, 2014, p. 171-182; C. CASONATO, *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, in *BioLaw Journal – Rivista di BioDiritto*, Special Issue, 2, 2019, p. 718 ss.

e delle finalità di un trattamento, e la complessità dei documenti con cui ciò accade rendono del tutto irrealistico che questi abbia la minima consapevolezza di quanto avviene coi suoi dati[335]. La situazione non potrà che ripetersi inalterata, sempre più frequentemente e in quasi ogni luogo, con la diffusione dell'internet delle cose. In più, stato più volte dimostrato che l'intelligenza artificiale, e in particolare alcune applicazioni del *machine learning*, rendono particolarmente problematiche le attività di pseudonimizzazione e anonimizzazione dei dati[336], rendendo necessario implementare strategie di crittografia sempre più dispendiose e complesse per evitare la reidentificazione[337]. Nemmeno quando il dato non può dirsi più dirsi personale, dunque, il contesto appare del tutto sicuro.

È chiaro che quanto appena esposto pone questioni inedite dal punto di vista della riservatezza, che si differenziano da quelle analizzate in precedenza per l'impossibilità di prevedere quali informazioni diffuse online potrebbero portare a rendere noti fatti, gusti o preferenze che preferiremmo tenere privati, chi entrerà in possesso di tali informazioni e per quali finalità le utilizzerà. Inoltre, le strategie basate sull'idea di architettura della scelta viste in precedenza, combinate con tecniche di analisi dei dati finalizzate alla profilazione, fanno ormai parte a pieno titolo del marketing commerciale e della comunicazione politica. Milioni di individui sono esposti, su *social network*, motori di ricerca, siti di *e-commerce*, piattaforme *streaming* e *app* di messaggistica, a pubblicità, suggerimenti e consigli in merito a prodotti da acquistare, contenuti di

[335] Significativo, sul punto, lo studio di A. M. MCDONALD, L. F. CRANOR, *The Cost of Reading Privacy Policies*, in *I/S: A Journal of Law and Policy for the Information Society*, 2008, p. 543-564 chestimava una perditapotenziale di 781 miliardi di dollari per il PIL annual statunitense, se gliutentiavesseroletto*veramente*tutte le privacy policy con cui si interfacciano online; riguardo all'eccessiva complessità di tali documenti, tale da renderli inaccessibili pressochè alla totalità dei consociati, cfr. E. SHERMAN, *Privacy policies are great – For PhDs*, in *CBS News*, 4 Sept. 2018, https://cbsn.ws/3h37Bt3; B. W. SCHERMER, B. CUSTERS, S. VAN DER HOF, *The Crisis of Consent cit.*, p. 10 ss; A. I. ANTON, J. B. EARP, Q. HE, W. STUFFLEBAUM, D. BOLCHINI, C. JENSEN, *Financial Privacy Policies and the Need for Standardization*, in *IEEE Security & Privacy*, 36, 2004, p. 42-44.

[336] L'art. 4 par. 1 n. 5 GDPR definisce pseudonimizzazione: «il trattamento dei dati personali in modo tale che i dati personali non possano più essere attribuiti a un interessato specifico senza l'utilizzo di informazioni aggiuntive, a condizione che tali informazioni aggiuntive siano conservate separatamente e soggette a misure tecniche e organizzative intese a garantire che tali dati personali non siano attribuiti a una persona fisica identificata o identificabile». La definizione di dati anonimi, ai quali, com'è noto, non si applica la disciplina del GDPR (al contrario che nel caso della pseudonimizzazione, nel quale il trattamento rimane disciplinato dal Regolamento), è enunciata al Considerando 26: «I principi di protezione dei dati non dovrebbero pertanto applicarsi a informazioni anonime, vale a dire informazioni che non si riferiscono a una persona fisica identificata o identificabile o a dati personali resi sufficientemente anonimi da impedire o da non consentire più l'identificazione dell'interessato. Il presente regolamento non si applica pertanto al trattamento di tali informazioni anonime,anche per finalità statistiche o di ricerca».

[337] Cfr. ad es. L. SWEENEY, M. VON LOEWENFELDT, M. PERRY, *Saying it's anonymous doesn't make it so: re-identification of "anonymized" law school data*, in *Technology Science*, November 12, 2018,https://techscience.org/a/2018111301/ (7 marzo 2021); L. SWEENEY, *Only you, your doctor and many others may know*, in *Technology Science*, September 28, 2015, https://techscience.org/a/2015092903/ (7 marzo 2021); A. J. COHEN, *New guarantees for cryptographic circuits and data anonymization*, Cambridge, 2019, p. 235; A. ANTONIOU, G. DOSSENA, J. MACMILLAN, S. HAMBLIN, D. CLIFTON, P. PETRONE, *Assessing the risk of re-identification arising from an attack on anonymized data*, 2022, arXiv:2203.16921.

cui usufruire, opinioni da condividere e comportamenti da tenere[338]. Il già citato CassSunstein ha chiamato questi luoghi digitali personalizzati *echo chambers*, additandoli come una delle cause principali della crescente, e a suo giudizio in larga parte irrazionale, polarizzazione della politica americana[339]. In realtà, tale sovraesposizione a contenuti personalizzati riguarda ormai i più vari ambiti dell'esistenza e, con la modifica dell'esperienza della navigazione internet che si prevede conseguirà alla diffusione dell'*IoT*, pare destinata ad investirla completamente.

La continua, artificialeeterodirezione dei desideri individuali in esame ha ovvie ripercussioni sulle possibilità di definire ed evolvere sé stessi, come detto al cuore della protezione dell'identità individuale. Il rischio, infatti, è generare una sorta di perpetuo *bias*di conferma di gusti, interessi ed opinioni, modellati su un profilo costruito, in ultima analisi, in funzione degli interessi commerciali, ideologici o di altra natura degli operatori privati che l'hanno messo a punto e non di quelli del singolo. Pare un colpo mortale per quello che è stato definito il "diritto alla discontinuità", ossia a modificare gusti, interessi, opinioni e comportamenti[340]. Da questo punto di vista, l'eventuale diffusione di sistemi di credito sociale non potrebbe che rappresentare un'ulteriore, e ancor più penetrante, intrusione nella sfera intima della persona. Alle possibilità di influenzare scelte e preferenze dischiuse dalla vista combinazione di tecniche di *nudging*e*machine learning*applicate ai dati personali si aggiungerebbe la pressione sociale generata dal timore delle sanzioni connesse ai comportamenti considerati scorretti, e dalla consapevolezza della costante sorveglianza da parte del sistema. Ciò porterebbe a un'interferenza di intensità del tutto inedita nella coscienza individuale e nella libertà di autodefinirsi e determinarsi del singolo, generando un'intrusionenelforo interno dell'individuo inimmaginabile anche per lo stato totalitario.

4. Ipotesi di regolazione e prospettive *de iure condendo* per una protezione efficace dei diritti riguardanti la sfera dell'identità nell'epoca dell'intelligenza artificiale

Alla luce di quanto affermato sinora, dunque, è evidente l'insufficienza dell'attuale quadro giuridico per la tutela dei diritti attinenti al libero sviluppo della personalità individuale. Come vedremo, è solo la prima di numerose sfide che l'intelligenza artificiale pone al diritto, evidenziando la necessità di innovare le norme esistenti, al fine di dare nuova protezione ai diritti fondamentali dell'individuo. Un eventuale intervento riparatore dovrebbe riguardare, in primo luogo, le norme

[338] Tra i molti contributi già citati, si rimanda in particolare a C. MELE, T. RUSSO SPENA, V. KAARTEMO, M. L. MARZULLO, *Smart nudging: how cognitive technologies enable choice architectures cit.*; K. YEUNG, *"Hypernudge": Big Data as a mode of regulation by design cit.*

[339] Cfr. C. SUNSTEIN, *#Republic.com: divided democracy in the age of social media*, Princeton, 2017.

[340] Così C. CASONATO, relazione *The Rise of New (and old) Rights in the Age of AI*, presso l'incontro inaugurale del corso *Constitutional Law of Technologies*, Firenze, 6 ottobre 2021; v. anche C. M. REALE, M. TOMASI, *Libertà d'espressione, nuovi media e intelligenza artificiale: la ricerca di un nuovo equilibrio nell'ecosistema costituzionale*, in *DPCE online*, 51, 1, 2022, p. 331.

sulla protezione dei dati personali, nelle quali, a prescindere dall'ordinamento di riferimento, il consenso informato del soggetto interessato riveste un ruolo fondamentale per la liceità di numerosi trattamenti[341]. Si tratta di un modello che, come abbiamo visto, si risolve in molti casi in una negazione delle tutele per le quali è nato[342]. Al fine di dare nuovo valore alla volontà dell'individuo, pare auspicabile un intervento in due direzioni. In primo luogo, limitando la possibilità degli operatori di internet di condizionare l'accesso a determinati servizi alla cessione da parte dell'utente di dati personali di per sé non necessari. Tale condotta contrattuale dovrebbe essere permessa solo quando siano presenti apposite garanzie per i diritti degli interessati,eappaia, in generale, ragionevole e proporzionata rispetto al contesto di riferimento, specialmente quando sia presente l'utilizzo di intelligenza artificiale particolarmente avanzata. Ai servizi della società dell'informazione, dunque, dovrebbe divenire possibile accedere acconsentendo unicamente ai trattamenti di dati necessari al loro funzionamento, e ulteriori finalità dovrebbero, in modo maggiore rispetto a ora, essere approvate caso per caso. Da questo punto di vista, alcuni recenti interventi intrapresi, a vario titolo, a livello europeo, paiono andare nella giusta direzione, come i limiti imposti all'utilizzo di c.d. *cookie wall*[343]. Si tratterebbe, in ogni caso, di un intervento da condurre con cautela, posti i limiti che ne deriverebbero per la libertà contrattuale, peraltro in analogia, almeno dal punto di vista europeo, con altri settori dell'ordinamento caratterizzati da una marcata asimmetria tra le parti (si pensi, ad esempio, ai contratti del consumatore). Andrà tenuto in considerazione, inoltre, il possibile impatto sulla gratuità di taluni servizi divenuti fondamentali per la società dell'informazione (come i *social network*) i cui ricavi derivano principalmente

[341] Si vedano anche le prospettive *de jure condendo*, di varia natura, di K. ISHII, *Comparative legal study on privacy and personal data protection for robots equipped with artificial intelligence: looking at functional and technological aspects*, in *AI & Society*, 34, 2019, p. 509-533, DOI 10.1007/s00146-017-0758-8; OFFICE FOR THE PRIVACY COMMISSIONER OF CANADA, *Consent and privacy. A discussion paper exploring potential enhancements to consent under the Personal Information Protection and Electronic Documents Act*, 2016, https://www.priv.gc.ca/media/1806/consent_201605_e.pdf (11marzo 2022); G. BUTTARELLI, *A smart approach: counteract the bias in artificial intelligence*, 8 novembre 2016, https://bit.ly/3JxjWUJ (1marzo 2022); K. MANHEIM, L. KAPLAN, *Artificial intelligence: risks to privacy and democracy*, p. 160 ss.; C. TUCKER, *Privacy, algorithms and artificial intelligence*, in *The Economics of Artificial Intelligence. an Agenda*, Chicago, 2019, p. 423-437; T.E. FROSINI, *La privacy nell'era dell'intelligenza artificiale*, in *DPCE Online*, 51, 1, 2022.

[342] V. ancora D. J. SOLOVE, *Privacy self-management cit.*; B. W. SCHERMER, B. CUSTERS, S. VON DERHOF, *The crisis of consent cit.*; C. CASONATO, *Costituzione e intelligenza artificiale cit.* e il resto dei contributi citati *supra*, n. 334.

[343] Con l'espressione *cookie wall* si indica una tecnica utilizzata dai siti web per impedire l'accesso agli utenti che non acconsentano a tutti i cookie e tracker presenti nella pagina. La pratica è stata dichiarata incompatibile con gli standard europei in materia di protezione dei dati personali dall'European Data Protection Board, poiché inibisce la formazione di un consenso genuino. Anche il Garante dei dati personali italiano ha, di recente, condiviso questa soluzione, facendola salva solo qualora il sito che impedisca l'accesso senza l'approvazione dei *cookie* garantisca comunque all'utente la possibilità di usufruire di un servizio equivalente. Questa soluzione è condivisa anche dal c.d. nuovo Regolamento ePrivacy, in fase di studio di fronte alle autorità europee e destinato a sostituire, nelle intenzioni di quest'ultime, la Dir. 2002/58/CE *relativa al trattamento dei dati personali e alla tutela della vita privata nel settore delle comunicazioni elettroniche* (nota, appunto, come Direttiva ePrivacy). Cfr. EDPB, *Guidelines 05/2020 on consent under Regulation 2016/679*, 4 maggio 2020, https://bit.ly/3xhbYN4 (12 marzo 2022); GPDP, *Linee guida cookies e altri strumenti di tracciamento*, 10 giugno 2021, https://bit.ly/3jywCju (12 marzo 2022); N. FABIANO, *ePrivacy, a che punto siamo? Ecco lo stato dell'arte*, in *Agenda digitale*, 1 marzo 2022 (12 marzo 2022).

dall'utilizzo di dati personali per finalità pubblicitarie. Non paiono, però, esigenze di fronte alle quali astenersi da un intervento proporzionato a tutela dei diritti e della libertà del singolo, e l'attuale entità dei ricavi delle società coinvolte non porta a temere per la loro permanenza sul mercato[344].

In secondo luogo, dovrebbe favorirsi lo sviluppo di una nuova cultura dei dati personali, che migliori la consapevolezza delle possibilità e dei rischi connessi al loro trattamento, in modo che sia le varie informative in materia di cui ogni individuo prende visione ogni giorno, che le eventuali richieste di consenso, possano finalmente raggiungere il loro scopo. Parallelamente, andrebbe favorito lo sviluppo di centri d'interesse col compito di rappresentare, a livello superindividuale, le prerogative degli interessati al trattamento dei dati personali[345]. Tali formazioni sociali (spesso indicate con l'espressione inglese *data trust*) potrebbero negoziare in rappresentanza di segmenti anche molto ampi della popolazione i termini del trattamento con gli operatori della società dell'informazione, sul modello di quanto avviene tra sindacati e associazioni datoriali in materia di contrattazione collettiva. Le prerogative degli individui acquisirebbero, così, un peso maggiore, e verrebbe rotto l'attuale schema di funzionamento del mercato, che vede contratti, informative e ogni altro documento relativo al trattamento predisposti unicamente dal proponente, spesso consistente in una piattaforma multinazionale, e rapidamente approvato, senza possibilità di modifica, dal singolo interessato. È ipotizzabile, inoltre, lo sviluppo di strategie affinché il singolo membro del *trust* venga rappresentato dal gruppo al momento di autorizzare o meno, al livello individuale, determinati trattamenti, sulla base di un insieme di valori e interessi generali stabilito al momento dell'adesione, migliorando, in tal modo, la consapevolezza di tali autorizzazioni, oggi pressoché nulla.

Parallelamente alla modifica di questi aspetti delle norme sulla protezione dei dati personali, dovrebbero introdursi alcune garanzie normative rivolte esplicitamente alla regolazione dell'intelligenza artificiale. Sembra auspicabile, prima di tutto, un'integrazione tra i due settori

[344] I ricavi delle *big tech*, infatti, non hanno fatto che aumentare nell'ultimo decennio (impennandosi, peraltro, in coincidenza dell'epidemia di Covid-19) e garantiscono alle aziende in questione margini di profitto di gran lunga più ampi di quelli degli operatori di ogni altro settore, cfr. F. ZANDT, *Big tech keeps getting bigger*, in *Statista*, 29 ottobre 2019, https://www.statista.com/chart/21584/gafam-revenue-growth/ (11 marzo 2022); O. WALLACH, *How big tech makes their billions*, in *Visual capitalist*, https://www.visualcapitalist.com/how-big-tech-makes-their-billions-2020/ (1 marzo 2022); C. GARTENBERG, *Big tech's 2021 earningswere off the chart*, The Verge, 11 febbraio 2022; S. OVIDE, *How big tech won the pandemic*, The New York Times, 30 aprile 2021 e *Big tech has outgrown this planet*, The New York Times, 29 luglio 2021.

[345] Sul punto, cfr. ad es. G. ZARKADIS, *"Data trusts" could be the key to better AI*, in *Harvard Business Review*, 10 Nov. 2020; K. HOUSER, J. W. BAGBY, *The data trust solution to data sharingproblems*, in *Vanderbilt Journal of Entertainment & Technology Law*, 2022, http://dx.doi.org/10.2139/ssrn.4050593; L. TREMOLADA, *Cosa sono i data trust e perchè possono aiutare la privacy e la società civile*, il Sole 24 Ore, 30 luglio 2021; S. DELACROIX, N. D. LAWRENCE, *Bottom-up data trusts: disturbing the "onesizefitsall" approach to data governance*, in *International Data Privacy Law*, 9, 4, 2019, p. 236-252; R. DI GIOACCHINO, F. STOLFI, *Data trust per un uso equo dei dati: un approccio contro lo strapotere delle Big Tech*, Agenda Digitale, 7 luglio 2021, https://bit.ly/3DYEwfn (9 marzo 2022).

normativi, con il riconoscimento di un particolare livello di rischio, da mitigare con opportune garanzie, ai trattamenti di dati svolti con alcune tecnologie basate sull'intelligenza artificiale, come l'applicazione del *machine learning* a grandi *dataset*[346]. A riguardo, il GDPR adotta già un approccio basato sul rischio, prevedendo, per i trattamenti in cui questo appaia maggiore, un apprezzabile insieme di garanzie procedurali e sostanziali (come lo svolgimento di valutazioni d'impatto del trattamento)[347] a tutela dei diritti dell'interessato, sotto il controllo delle autorità garanti[348]. Tale impostazione, pur muovendosi nella giusta direzione, non pare del tutto sufficiente, poiché non menziona esplicitamente, come pare auspicabile, il coinvolgimento di talune tecnologie intelligenti tra i presupposti per identificare come rischioso un trattamento.

Infine, sarebbe bene riconoscere che alcune applicazioni dell'intelligenza artificiale presentano rischi per i diritti fondamentali che non appaiono facilmente mitigabili attraverso la regolazione e dovrebbero, dunque, essere messe al bando. Si tratta di un tema che accompagnerà l'intero lavoro, ripresentandosi per ciascuno dei diritti che verranno analizzati. Riguardo ai diritti attinenti alla sfera dell'identità, le tecnologie che paiono porre rischi inaccettabili sono soprattutto quelle che fanno uso delle tecniche più avanzate di condizionamento del comportamento degli individui e dei sistemi di credito sociale. Nella totale assenza, sul piano globale, di strumenti normativi in materia, la Proposta di Regolamento dell'Unione Europea in materia di intelligenza artificiale dimostra una buona consapevolezza del problema, mettendo al bando le tecnologie intelligenti volte alla distorsione subliminale del comportamento degli individui, qualora sfruttino debolezze specifiche o siano volte a ottenere effetti negativi per quest'ultimi o terzi[349]. Maggiori perplessità suscita, invece,

[346] L'art. 35 del GDPR, in materia di valutazione d'impatto sulla protezione dei dati, fa solo un generico riferimento, tra gli indicatori di possibile rischio elevato del trattamento, alla circostanza che questo sia svolto con «l'uso di nuove tecnologie». Il Cons. 75, invece, menziona, allo stesso scopo, il fatto che «il trattamento riguarda una notevole quantità di dati personali e un vasto numero di interessati». Le stesse parole *intelligenza artificiale* non compaiono mai nel testo del Regolamento. Cfr. anche le considerazioni di L. MITROU, *Data protection, artificial intelligence and cognitive services. Is the General Data Protection Regulation (GDPR) Artificial Intelligence-proof?*, 2018, https://dx.doi.org/10.2139/ssrn.3386914; P. SCHWARTZ, *Risk and high-risk: Walking the GDPR tighrope*, in *IAPP*, 29 marzo 2016, https://bit.ly/37TJVIS (17 marzo 2022) ; R. GELLERT, *The role of the risk-based approach in the General Data Protection Regulation and in the European Commission's proposed Artificial Intelligence Act: business as usual?*, in *Journal of Ethics and Legal Technologies*, 3, 2, 2021, p. 15-33.

[347] Cfr. gli artt. 35 e 36 del Regolamento.

[348] Sul punto, sivedano, da varipunti di vista, R. GELLERT, *The risk-based approach to data protection*, New York, 2020 e *Understanding the notion of risk in the General Data Protection Regulation*, in *Computer Law & Security Review*, 34, 2, 2018, p. 279-288; M. E. GONCALVES, *The risk-based approach under the new EU data protection regulation: a critical perspective*, in *Journal of risk research*, 23, 2, 2020, p. 139-152; G. MALDOFF, *The risk-based approach in the GDPR: interpretation and implications*, IAPP – White Paper, 2016, https://bit.ly/3uQKr3x (17 marzo 2022); P. ZANELLATI, *GDPR, l'approccio privacy basatosulrischio: come funziona e le misure da adottare*, in *Cybersecurity360*, 4 dicembre 2020, https://bit.ly/36mornF (17 marzo 2022).

[349] È l'art. 5 della Proposta di Regolamento ad elencare le applicazioni dell'intelligenza artificiale vietate nell'ordinamento europeo. Nello specifico, le lett. a e b del par. 1 bandiscono: «a) l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che utilizza tecniche subliminali che agiscono senza che una persona ne sia consapevole al fine di distorcerne materialmente il comportamento in un modo che provochi o possa provocare a tale persona o a un'altra persona un danno fisico o psicologico; b) l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che sfrutta le vulnerabilità di uno specifico gruppo di persone, dovute all'età o alla disabilità fisica o

la norma in materia di sistemi di credito sociale e altri strumenti di valutazione dell'affidabilità delle persone fisiche, messi al bando unicamente quando comportino un trattamento sfavorevole della persona in contesti diversi da quello in cui i dati alla base del funzionamento del sistema sono stati raccolti, o comunque sproporzionato rispetto alle condotte tenute[350]. Questa attuale versione, infatti, non pare idonea a mettere al riparo gli individui dal deterioramento della tutela di numerosi diritti fondamentali e dalla diminuzione degli spazi democratici che possono conseguire alla diffusione di sistemi di credito sociale[351]. Che dire, ad esempio, di un sistema di credito sociale "completo", in grado di valutare il comportamento del cittadino in ogni ambito dell'esistenza, e i cui risultati, quindi, non potrebbero mai essere fuori contesto rispetto ai dati raccolti, paradossalmente rispettando, all'apparenza, i limiti imposti dalla bozza di Regolamento europeo? Pare necessario, dunque, rimettere mano alla sua formulazione, in modo da escludere in modo più chiaro il realizzarsi di scenari che appaiono, francamente, distopici. Allo stesso tempo, si deve evidenziare che strumenti valutativi del comportamento da cui derivano circoscritte conseguenze in positivo o in negativo esistono già in numerosi servizi della c.d. *gig economy* (si pensi alle app di *car sharing* o di alloggio turistico basate sulle recensioni degli utenti) senza che da essi derivi particolare allarme sociale e che hanno, anzi, acquisito un apprezzabile valore economico. Probabilmente, è stata la consapevolezza di dover preservare tali servizi a spingere all'attuale formulazione della norma che mette al bando determinati sistemi di credito sociale. Pare possibile e auspicabile, però, individuare una formula che distingua nettamente tra situazioni obiettivamente diverse (e, forse, inopinatamente accostate proprio dall'attuale versione del Regolamento) e garantisca la funzionalità dei servizi della *gig economy* proteggendo, al contempo, i diritti fondamentali dell'individuo da applicazioni tecnologiche che richiamano lo stato totalitario.

---

mentale, al fine di distorcere materialmente il comportamento di una persona che appartiene a tale gruppo in un modo che provochi o possa provocare a tale persona o a un'altra persona un danno fisico o psicologico».

[350]La disposizione interessata è l'art. 5 par. 1 lett. c della Proposta di Regolamento, che vieta: «l'immissione sul mercato, la messa in servizio o l'uso di sistemi di IA da parte delle autorità pubbliche o per loro conto ai fini della valutazione o della classificazione dell'affidabilità delle persone fisiche per un determinato periodo di tempo sulla base del loro comportamento sociale o di caratteristiche personali o della personalità note o previste, in cui il punteggio sociale così ottenuto comporti il verificarsi di uno o di entrambi i seguenti scenari: i) un trattamento pregiudizievole o sfavorevole di determinate persone fisiche o di interi gruppi di persone fisiche in contesti sociali che non sono collegati ai contesti in cui i dati sono stati originariamente generati o raccolti; ii) un trattamento pregiudizievole o sfavorevole di determinate persone fisiche o di interi gruppi di persone fisiche che sia ingiustificato o sproporzionato rispetto al loro comportamento sociale o alla sua gravità».

[351]Si tratta di preoccupazioni in parte fatte proprie anche dal COMITATO ECONOMICO E SOCIALE EUROPEO, nel suo parere sulla Proposta di Regolamento adottato il 22 settembre 2021, in cui si legge (p. 3): «Il CESE ritiene che nell'UE non vi sia cittadinanza per un sistema che attribuisce un punteggio all'affidabilità dei cittadini europei sulla base del loro comportamento sociale o delle caratteristiche della loro personalità, quale che sia l'attore che assegna tale punteggio. Il CESE raccomanda di ampliare l'ambito di applicazione di questo divieto in modo da includervi il punteggio sociale da parte di organizzazioni private e autorità semipubbliche», https://bit.ly/3uP5vXZ (21 marzo 2022).

Nuove sfide per "vecchi diritti". Intelligenza artificiale e libera manifestazione del pensiero: la *content moderation* automatizzata dei *social media*

1. Il ruolo dei *social media* nella comunicazione contemporanea e le questioni in materia di libertà d'espressione poste dalla moderazione dei contenuti

Nei paragrafi precedenti, la diffusione dei servizi di internet interattivo (c.d. web 2.0)[352] è stata più volte menzionata tra i fattori determinanti crescenti vuoti di tutela nella protezione giuridica della sfera dell'identità. Ciò che ora preme evidenziare, invece, è come il loro effetto trasformativo sia stato ancora maggiore per quanto riguarda campo di applicazione, limiti e ambito di protezione del diritto alla libera manifestazione del pensiero. L'utilizzo crescente, in tali servizi, di tecnologie di intelligenza artificiale ha avuto un ruolo decisivo anche da questo punto di vista[353]. L'obiettivo dei prossimi paragrafi sarà, appunto, analizzare nel dettaglio le questioni aperte in materia di libertà d'espressione dall'intreccio di intelligenza artificiale e web 2.0.

É probabilmente superfluo rilevare il ruolo centrale che la libera manifestazione del pensiero riveste in ogni studio che si occupi dei diritti fondamentali dell'individuo. Ci si limiterà a ricordare come essa, garantendo la possibilità di comunicare all'esterno, diffondere e contestare opinioni e orientamenti ideologici, politici e religiosi, siauno dei perni fondanti di diversidei principi cardine della Costituzione italiana, come il principio personalista o il principio pluralista[354]. Non è un caso, dunque, che sia stata definita dalla Consulta, con un'espressione sin troppo nota, la «pietra angolare

[352] Come già chiarito (cfr. n. 258) il termine è stato coniato da D. DINUCCI, *Fragmented future cit.*, ed è diventato di uso comune a partire dalla prima edizione della *O'Rielly Web 2.0 Conference*, tenutasi a San Francisco nel 2004, https://bit.ly/3sqzOTv (30 maggio 2022).

[353] Per alcuni trattazioni generali del tema possono indicarsi, da diverse prospettive e tra i molti, C. SUNSTEIN, *Republic.com*, Princeton, 2001; *Republic.com 2.0*, Princeton, 2007; *#Republic.com: divideddemocracy in the age of social media*, Princeton, 2017; J. M. BALKIN, *Free speech in the algorithmic society: big data, private governance, and new schoolspeechregulation*, in *U.C. Davis Law Review*, 51, 3, 2018, p. 1149-1210; G. PITRUZZELLA, O. POLLICINO, S. QUINTARELLI, *Parole e potere. Libertàd'espressione, hate speech e fake news*, Milano, 2017; E. PARISIER, *The filter bubble: what the internet is hiding from you*, New York, 2011; D. A. GALLAGHER, *Free speech on the line: modern technology and the First Amendment*, in *CommLaw Conspectus: Journal of Communications Law & Policy*, 3, 2, 1995, p. 197-206; J. KOSSEFF, *The twenty-six words that created the internet*, Ithaca, 2019; R. GORWA, R. BINNS, C. KATZENBACH, *Algorithmic content moderation: Technical and political challenges in the automation of platform governance*, in *Big Data & Society*, 7, 1, 2020; E. LLLANSÒ, J. VON HOBOKEN, P. LEERSSEN, J. HARAMBAM, *Artificial intelligence, content moderation and freedom of expression*, Working Paper – Transatlantic Working Group on content moderation online and freedom of expression, 2019; CAMBRIDGE CONSULTANTS, *Report produced on behalf of Ofcom - Use of AI in online content moderation*, 2019, https://bit.ly/3bdRqX1 (30 aprile 2022); S. J. BRISON, K. GELBER, *Free speech in the digital age*, Oxford, 2019; C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Diritto pubblico comparato ed europeo*, numero speciale, 2019, p. 110 ss. e *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, in *BioLaw Journal – Rivista di Biodiritto*, Special Issue 2, 2019, p. 713 ss; M. FASAN, *Intelligenza artificiale e pluralismo: uso delle tecniche di profilazione nello spazio pubblico democratico*, in *BioLaw Journal – Rivista di Biodiritto*, 1, 2019, p. 107 ss.

[354] Si rimanda, in generale e tra i molti, a A. PACE, M. MANETTI, *La libertà di manifestazione del pensiero. Art. 21*, in G. BRANCA, A. PIZZORUSSO, *Commentario della Costituzione*, XI, Bologna, 2006; C. ESPOSITO, *La libertà di manifestazione del pensiero nell'ordinamento italiano*, Milano, 1958; P. BARILE, *Libertà di manifestazione del pensiero*, Milano, 1975 e *Diritti dell'uomo e libertà fondamentali*, Bologna, 1984, p. 210 ss.

dell'ordinamento democratico»[355]. Sul piano comparato, inoltre, la libertà d'espressione offre spunti di riflessione di particolare interesse, viste le profonde differenze che ne caratterizzanola concezione tra i distinti ordinamenti, in particolare riguardo allo spazio ad essa riservato nei rapporti tra privati[356].

Tra i diversi servizi di internet interattivo vengono in rilevo, ai fini di quest'analisi, in particolare i c.d. *social media*, il cui funzionamento si basa in larghissima parte sulla generazione di contenuti da parte degli utenti stessi. Sebbene non sia semplice individuare una precisa data di nascita del fenomeno, è possibile affermare che i primi tentativi di dar vita a piattaforme online simili ai moderni *social network* risalgano alla fine degli anni 90' del XX secolo[357]. Le multinazionali che oggi dominano il mercato, invece, sono state fondate in larga parte nel primo decennio del XXI secolo, per acquisire la loro posizione di predominio in quello successivo[358].Pur nell'estrema diversità di funzionamento che li caratterizza, ciascuno di questi strumenti fa della possibilità per gli utenti di condividere contenuti (tipicamente testi, immagini, audio e video) l'elemento centrale del proprio servizio. A ciò si aggiunge la possibilità di costruire relazioni di vario genere, con la possibilità di certificare rapporti di "amicizia" tra diversi profili personali, usare servizi di messaggistica o manifestare apprezzamento e interesse (ad es. attraverso i c.d. *like*) [359]. Chiaramente, questi nuovi contesti digitali hanno presto attirato l'attenzione dei media tradizionali, che vi hanno visto, allo stesso tempo, una pericolosa forma di concorrenza, specialmente sul terreno dell'informazione, e l'opportunità di raggiungere segmenti del pubblico non raggiungibili con gli

[355] L'espressione compare in Corte cost. sent. n. 84 del 1969. In senso analogo, Corte cost. sent. n. 126 del 1985 che definisce il diritto di cui all'art. 21 «cardine di democrazia nell'ordinamento generale», Corte cost.sent. n. 11 del 1968, che lo descrive come «coessenziale al regime di libertà garantito dalla Costituzione», Corte cost. n. 126 del 1985, che evidenzia «la rilevanza centrale […] che la libertà di manifestazione del pensiero, anche e soprattutto in forma collettiva, assume ai fini dell'attuazione del principio democratico»

[356] Cfr., tra i molti, M. ROSENFELD, A. SAJO, *Spreading liberal constitutionalism: an inquiry into the fate of free speech rights in new democracies*, in S. CHOUDRY (A CURA DI), *The migration of constitutional ideas*, Cambridge, 2007; O. POLLICINO, *La prospettiva costituzionale sulla libertà d'espressione nell'era di internet*, in G. PITRUZZELLA, O. POLLICINO, S. QUINTARELLI, *Parole e potere cit.*; O. POLLICINO, *L'efficacia orizzontale dei diritti fondamentali previsti dalla Carta. La giurisprudenza della Corte di giustizia in materia di digital privacy come osservatorio privilegiato*, in *MediaLaws*, 3, 2018; O. POLLICINO, M. BASSINI, *Free speech, defamation and the limits to freedom of expression in the EU: a comparative analysis*, in A. SAVIN, J. TRZASKOWSKI (A CURA DI), *Research Handbook On EU Internet Law*, Cheltenham-Northampton, 2014, p. 508 ss.; M. OROFINO, *La libertà di espressione tra Costituzione e Carte europee dei diritti*, Torino, 2014; C.R. SUNSTEIN, *Democracy and the Problem of Free Speech*, New York, 1995.

[357] Il primo *social network* è in genere considerate *SixDegrees*, fondato dal newyorkese Andrew Weinreichnel 1997, cfr. D.M. BOYD, N.B. ELLISON, *Social network sites: definition, history, and scholarship*, in *Journal of computer-mediated communication*, 13, 1, 11, http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html (30 aprile 2022)

[358] *Facebook*, ad esempio, è stato fondato nel 2004 (al pari di *Netlog*, il principale *social network* nato in Europa, chiuso nel 2014), Twitter nel 2006, la prima versione di Linkedin risale addirittura al 2002. VKontakte, il principale socialrusso, è stato fondato nel 2006; il cinese QZone (oggi soppiantato dai social network integrati nell'appWeChat, nata nel 2011) nel 2005.

[359] Secondo la definizionedellecitateD.M. BOYD, N.B. ELLISON, *Social network sites cit*: «We define social network sites as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site».

strumenti comuni. Pressoché ogni giornale, rivista, canale televisivo o radiofonico, così, ha cominciato ad operare anche sui social media, al pari di aziende private, organizzazioni non governative, partiti politici, associazioni culturali e altri centri d'interesse, che vi hanno visto l'opportunità per lo sviluppo di nuove tipologie di pubblicità, proselitismo o propaganda. Parallelamente, sono nati operatori dell'informazione e commentatori, generalisti o specializzati nei campi più vari, che, con l'impiego di risorse molto contenute e generalmente non registrati con le procedure previste, in molti ordinamenti, per i media ufficiali, sono spesso stati in grado di raggiungere un seguito digitale composto da diverse migliaia di utenti. Sempre più spesso semplici individui – indicati come *influencer* – hanno raggiunto una popolarità tale sui *social media* da diventare *testimonial* pubblicitari appetibili per società multinazionali[360]. Un contenuto, diffuso da un singolo utente di un *social media*, può venire visualizzato da milioni di persone in poche ore grazie a continue condivisioni, diventando "virale". In breve, le piattaforme hanno acquisito, in pochi anni, un ruolo centrale nel mercato dell'informazione, e oggi non esiste un altro mezzo che permetta di raggiungere un numero tanto ampio di utenti, specialmente riguardo a talune fasce di popolazione meno legate ai media tradizionali[361]. I loro ricavi, come già detto, derivano in larga parte da introiti pubblicitari, posta la gratuità dei loro servizi, del resto strumentale al loro funzionamento, favorendo la presenza e l'interazione del maggior numero possibile di utenti.

Un modello di business basato sulla produzione e diffusione di contenuti generati o detenuti dagli utenti porta alla necessità di sviluppare forme di sorveglianza su quegli stessi contenuti (c.d. *content moderation*). In primo luogo, per evitare la diffusione di materiali vietati in varia misura in molti ordinamenti, spesso con norme di diritto penale (si pensi alla pedopornografia e ad alcune forme di discorsi d'odio o di istigazione a delinquere)[362] o soggetti a precisi limiti di legge (contenuti a sfondo erotico e sessuale, pubblicità di determinati prodotti la cui commercializzazione è soggetta a restrizione)[363]. In secondo luogo, per limitare la diffusione di materiale che, pur non apparendo *prima facie* vietato dalla legge, non sembri opportuno diffondere sulla piattaforma. Si tratta, in

[360] Tanto da suscitare in più occasioni l'interesse dell'*Economist*, cfr. *The rise of the influencer economy*, The Economist, 2 aprile 2022 e *The new rules of the "creator economy"*, The Economist, 8 maggio 2021.

[361] Cfr. in generale M. KENNEY, J. ZYSMAN, *The rise of platform economy*, in *Issues in science and technology*, 2016, p. 61-69; J. E. COHEN, *Law for the platform economy*, in *U.C. D. Law Review*, 51, 2018, p. 133 ss.; D.S. EVANS, R.L. SCHMALENSEE, *Matchmakers: the new economy of multisided platforms*, Harvard, 2016; A. GAWER, *Platforms, markets and innovation*, Cheltenham, 2010.

[362] Si tratta di criminalizzazioni, peraltro, spesso legate a impegni condivisi in seno alla comunità internazionale. Si pensi alla *Convenzione ONU sui diritti del fanciullo*, adottata a New York nel 1989, e in particolare al suo *Protocollo opzionale relativo alla vendita di minori, la prostituzione e la pornografia infantile*, risalente al 2000, o alla *Convenzione internazionale sull'eliminazione di ogni forma di discriminazione razziale*, adottata dall'Assemblea Generale delle Nazioni Unite nel 1965.

[363] In Italia, ad esempio, La L. 52 del 1983 vieta la propaganda pubblicitaria di ogni prodotto da fumo; la pubblicità di bevande alcoliche è sottoposta a limiti stringenti dalla L. n. 125 del 2001; il D.L. n. 87 del 2018 ha introdotto il divieto di pubblicità di giochi e scommesse, con l'esclusione di quelle con mera finalità informativa e descrittiva per favorire scelte di gioco consapevoli, indipendente dal media utilizzato (televisione, radio o, appunto, internet).

questo caso, di un'ampia "zona grigia" in cui possono rientrare contenuti particolarmente disturbanti (come immagini di vittime di delitti di sangue o di contesti bellici)o che pare opportuno limitare per ragioni di opportunità (si pensi alla proliferazione di notizie false). Se nel primo caso sottrarre i contenuti alla visione degli utenti pare corrispondere agli interessi commerciali delle piattaforme, non si può dire lo stesso con altrettanta sicurezza della limitazione della diffusione di notizie false, che non appaiono, di per sé, idonee a scoraggiare gli utenti dall'utilizzare la piattaforma. Anche per questo, il tema ha attirato l'attenzione della stampa e dell'opinione pubblica solo dopo che la massiccia diffusione di *fake news* sui *social media* è sembrata in grado di influenzare pesantemente alcuni avvenimenti della vita collettiva, come elezioni, referendum o mobilitazioni di massa[364]. Come si dirà, lo stesso atteggiamento delle piattaforme, in seguito a tali polemiche, non è stato lineare e varia molto da un operatore all'altro.

L'attività di moderazione dei contenuti generati o diffusi dagli utenti pone, chiaramente, interrogativi interessanti e di difficile soluzione, posta la natura sostanzialmente censoria di molte di tali decisioni, la delicatezza del bilanciamento di interessi che talvolta richiedono e la natura squisitamente privata dei soggetti che le mettono in atto[365]. Prima di affrontare queste questioni,

[364]Polemiche, in particolare, sono sorte dalle rivelazioni connesse alle attività della società di consulenza Cambridge Analyitica e dalla manipolazione dell'opinione pubblica attraverso la diffusione di notizie false sui social network che ha caratterizzato il referendum sull'uscita del Regno Unito dall'Unione Europea e le elezioni americane del 2016. Cfr. H. MARSHALL, A. DRIESCHOVA, *Post-truth politics in the UK's Brexit referendum*, in *New Perspectives*, 26, 3, 2018, p. 89–106; R. KÜBLER, K. PAUWELS, K. MANKE, *How Social Media Drove the 2016 US Presidential Election: A Longitudinal Topic and Platform Analysis*, Rochester Social Science Research Network, luglio 2020, https://papers.ssrn.com/abstract=3661846 (2 maggio 2022); Y. TSFATI, H.G. BOOMGAARDEN, J. STRÖMBÄCK, R. VLIEGENTHART, A. DAMSTRA, E. LINDGREN, *Causes and consequences of mainstream media dissemination of fake news: literature review and synthesis*, in *Annals of the International Communication Association*, 44, 2, 2020, p. 157-173; A. R. DOSHI, S. RAGHAVAN, R. WEISS, E. PETITT, *How the supply of fake news affected consumer behavior during the 2016 US election*, 2018, http://dx.doi.org/10.2139/ssrn.3093397 (2 maggio 2022).

[365] Cfr., in via estremamente generale e da vari punti di vista, R. PERRONE, *Fake news e libertà di manifestazione del pensiero: brevi coordinate in tema di tutela costituzionale del falso*, in *Nomos-Le attualità del diritto*, 2, 2018, p. 25 ss; C. MELZI D'ERIL, *Fake news e responsabilità: paradigmi classici e tendenze incriminatrici*, in *MediaLaws – Rivista di diritto dei media*, 1, 2017, p. 60 ss; S. FOÀ, *Pubblici poteri e contrasto alle fake news. Verso l'effettività dei diritti aletici?*, in *Federalismi.it*, 11, 2020; M. MONTI, *La disinformazione online, la crisi del rapporto pubblico-esperti e il rischio della privatizzazione della censura*, in *Federalismi.it*, 11, 2020; *Privatizzazione della censura e internet platforms: la libertà di espressione e i nuovi censori dell'agorà digitale*, in *Rivista italiana di informatica e diritto*, 1, 2019, p. 35 ss.;F. DONATI, *Fake news e libertà d'informazione*, in *Medialaws – Rivista di diritto dei media*, 2, 2018, p. 445 ss.;C. MAGNANI, *Libertà d'informazione online e fake news: vera emergenza? Appunti sul contrasto alla disinformazione tra legislatori statali e politiche europee*, in *Forumdi Quaderni costituzionali – Rassegna*, 4, 2019, p. 16 ss.; P. MOURON, *Une future loi pour luttercontrelesfake news: lesdifficultés d'une définitionjuridique*, in *Revueeuropéennedesmédias et dunumérique*, 2018, 45, p. 66 ss.; FROSINI T.E., *Internet come ordinamento giuridico*, in *Percorsi costituzionali*, 1, 2014, p. 262 ss. e FROSINI T.E., *No news is fake news*, in DPCE, 4, 2017, p. 4 ss.; C. M. REALE; M. TOMASI, *Libertà d'espressione, nuovi media e intelligenza artificiale: la ricerca di un nuovo equilibrio nell'ecosistema costituzionale*, in *DPCE online*, 51, 1, 2022; T. GILLESPIE, *Custodians of the Internet: Platforms, Content Moderation, and the HiddenDecisionsthatShape Social Media*, New Haven, 2018; G. PITRUZZELLA, O. POLLICINO, S. QUINTARELLI, *Parole e potere cit.*; O. POLLICINO, M. BASSINI, *Free speech, defamation and the limits to freedom of expression cit.*; G. MARCHETTI, *Le fake news e il ruolo degli algoritmi*, in *MediaLaws–Rivista di diritto dei media*, 1, 2020, p. 29-35; A. LAURO, *Siamo tutti giornalisti? Appuntisullalibertà di informazionenell'era social*, in *MediaLaws – Rivista di dirittodei media*, 2, 2021, p. 1–24; J. M. BALKIN, *Free speech in the algorithmic society cit.*; D. A. GALLAGHER, *Free speech on the line cit.*; S. J. BRISON, K. GELBER, *Free speech in the digital age cit.*; R. GORWA, R.

però, è necessario analizzare quale sia il ruolo delle tecnologie basate sull'intelligenza artificiale nell'automazione su larga scala di tale attività di moderazione, dalla quale non pare potersi prescindere, visto il volume dei contenuti coinvolti. Un'applicazione dell'intelligenza artificiale che, com'è intuibile, pone questioni ulteriori oltre a quelle già accennate, in primo luogo dal punto di vista del diritto alla libera manifestazione del pensiero degli utenti dei *social media*.

2. L'intelligenza artificiale nella *content moderation*: cenni sulle principali tecnologie coinvolte nel filtro dei contenuti diffusi sui *social media*

Semplificando fino all'essenziale, il filtro dei contenuti su un *social media* si basa su due attività fondamentali: la ricerca di copie, identiche o con lievi modifiche, di materiale già noto (indicabile col termine inglese *matching*) e l'analisi di contenuti originali o comunque non noti, al fine di stabilire se appartengano a una categoria vietata (indicabile come *classification*)[366]. L'attività di *matching*viene in rilievo, in particolare, al momento di impedire la diffusione di immagini e video già illecitamente diffusi in precedenza. I sistemi più utilizzati dalle principali piattaforme si basano sulla tecnica crittografica dell'*hashing*, consistente nell'assegnare, con un'apposita funzione, un valore numerico (*hash*) a un determinato contenuto[367]. Ciascuna copia di tale contenuto si vedrà, da quel momento in poi, attribuire lo stesso valore dalla funzione, venendo immediatamente riconosciuta e filtrata dal sistema. La tecnica si basa sull'unicità di ogni codice *hash*, associato a un singolo contenuto, con scarsissime possibilità di sovrapposizione. Ciò, però, porta a un'estrema sensibilità del sistema anche a modifiche minime dei contenuti analizzati, che non vengono riconosciuti dalla funzione. Per questa ragione, sono state sviluppate una grande varietà di tecniche in grado di identificare contenuti che appaiano simili nei loro elementi fondamentali. Tali sistemi, utilissimi nell'attività di *content moderation*, rinunciano al principio crittografico dell'unicità della corrispondenza tra codice *hash* e singolo contenuto, assegnando valori simili a materiali simili e svolgendo l'attività di moderazione in base a tali similitudini[368]. Deve segnalarsi che diverse tra le principali piattaforme hanno intrapreso imponenti sforzi congiunti riguardo a questa tecnologia, specialmente in materia di contrasto al terrorismo. Nel 2017 Facebook, Twitter, Google e Microsoft

---

BINNS, C. KATZENBACH, *Algorithmic content moderation cit.*; E. LLANSÒ, J. VON HOBOKEN, P. LEERSSEN, J. HARAMBAM, *Artificial intelligence, content moderation and freedom of expression cit.*

[366] Cfr. R. GORWA, R. BINNS, C. KATZENBACH, *Algorithmic content moderation cit.*, p. 4 ss.; N. DUARTE, E. LLANSO, A. LOUP, *Mixed messages? The limits of automated social media content analysis*, Report fo the Center for Democracy & Technology, 2017, p. 7 ss.; E. ENGSTROM, N. FEAMSTER, *The limits of filtering: alook at the functionality and shortcomings of content detection tools*, Report - Engine, 2017, http://www.engine.is/the-limits-of-filtering/, p. 11 ss.

[367] A. J. MENEZES, P. C. VON OORSCHOT, S. A. VANSTONE, *Handbook of applied cryptography*, 1997, p. 321 ss.; *Hashing*, Techopedia, 2021, https://bit.ly/3mc9mJv (3 maggio 2022); R. GORWA, R. BINNS, C. KATZENBACH, *Algorithmic content moderation cit.*, p. 4-5; E. ENGSTROM, N. FEAMSTER, *The limits of filtering cit.*, p. 11 ss.

[368] Cfr. ancora R. GORWA, R. BINNS, C. KATZENBACH, *Algorithmic content moderation cit.*, p. 4-5; E. ENGSTROM, N. FEAMSTER, *The limits of filtering cit.*, p. 11 ss.

hanno dato vita al *Global Internet Forum to Counter Terrorism*, in seno al quale hanno dato vita a un *database* condiviso dei valori di *hash*di contenuti illeciti relativi ad attività terroristiche, oggi usato da 13 diverse piattaforme[369]. Questa iniziativa ha portato a risultati molto significativi, come il filtro automatico, in tempi molto brevi, di centinaia di diverse versioni, diffuse in varie piattaforme, del video dell'attentato alla moschea di Al Noor di Christchurch, in Nuova Zelanda, compiuto nel 2019, in diretta *streaming* su Facebook, da un estremista di destra dotato di *bodycam*, in cui hanno perso la vita più di cinquanta persone[370].

La moderazione di immagini e video nelle reti sociali, in ogni caso, non è affidata solamente alla tecnica dell'*hashing*, del resto strutturalmente inadatta a riconoscere e filtrare contenuti creati dallo stesso utente che li diffonda[371]. Le reti sociali fanno, contemporaneamente, ampio uso di sistemi di *computer vision*, basati su diverse tecnologie, generalmente coinvolgenti tecniche di intelligenza artificiale. Si tratta, come si dirà, di tecnologie che danno adito a forti discussioni, posto che la loro accuratezza – notevole, ma ancora imperfetta – potrebbe portare a moderare contenuti in misura maggiore al necessario.

Oltre a immagini e materiali audio e video, la *content moderation* delle piattaforme si svolge su un'enorme mole di contenuti testuali, generati o condivisi dagli utenti. Generalmente, l'attività automatizzata è svolta con l'ausilio di tecnologie intelligenti appartenenti al campo del *Natural Language Processing*, la branca dell'intelligenza artificiale che punta allo sviluppo di sistemi in grado di comprendere, tradurre e generare il linguaggio naturale[372]. Il filtro di contenuti in formato testuale si basava, in origine, sulla ricerca di determinate parole chiave – una delle prime applicazioni furono i sistemi *antispam* dei servizi di posta elettronica – oggi, invece, si basa su

[369] S. LARSON, *Tech giants bolster collaborative fight against terrorism*, CNN Business, 26 giugno 2017, https://cnn.it/3x5XRs9 (3maggio 2022). Per una panoramica delle attività del Global Internet Forum to CounterTerrorism, può consultarsi il sito ufficiale dell'iniziativa: https://gifct.org/ (3 maggio 2022).

[370] Le dichiarazioni ufficiali di Facebook in proposito parlavano dell'identificazione e rimozione di più di 800 versioni diverse del video in poche ore, cfr. *Update on new Zeland*, 18 marzo 2019, https://perma.cc/ZA85-2Y3X (3 maggio 2022); v. anche R. GORWA, R. BINNS, C. KATZENBACH, *Algorithmic content moderation cit.* Non sono mancate, comunque, narrazioni della stampa di segno in parte contrario, cfr. J. WAKEFIELD, *Christchurch shootings. Social media races to stop attack footage*, BBC News, 16 marzo 2019, https://www.bbc.com/news/technology-47583393 (3maggio 2022).

[371] Sono diffusi, in particolare, sistemi di moderazione dei contenuti basati sull'analisi dei c.d. metadati (informazioni che riguardano le caratteristiche di un file determinato, come il titolo, l'autore, o la lunghezza se si fa riferimento a una canzone). Si tratta di strumenti imprecisi e con un margine d'errore non trascurabile, posto che contenuti diversi possono avere metadati simili o identici, e che questi ultimi possono essere manipolati e criptati per evitare il riconoscimento. Altri strumenti, sempre più utilizzati, si basano sulla costruzione di un'"impronta digitale" del contenuto moderato ricavata da alcune caratteristiche (come determinate frequenze sonore ricorrenti in una canzone), e sembrano garantire prestazioni migliori di quelli operanti sui metadati o con tecniche di *hashing*. Il loro limite principale sta nella loro ristretta versatilità: un sistema di *fingerprinting* audio, ad esempio, non potrà essere utilizzato per l'analisi di immagini, per le quali sarà necessario lo sviluppo di ulteriori strumenti *ad hoc*, cfr. E. ENGSTROM, N. FEAMSTER, *The limits of filtering cit.*, p. 11 ss.

[372] Cfr. N. DUARTE, E. LLANSO, A. LOUP, *Mixed messages? The limits of automated social media contentanalysis cit.*, p. 7 ss; S. QUINTARELLI, *Content moderation: i rimedi tecnici*, in G. PITRUZZELLA, O. POLLICINO, S. QUINTARELLI, *Parole e potere cit.*, p. 110 ss.

tecnologie estremamente complesse, che combinano diversi approcci all'intelligenza artificiale e prendono in considerazione segmenti del discorso molto più ampi di una singola parola chiave, oltre a disporre di strategie per inferire la probabile funzione grammaticale di determinati elementi di una frase[373]. L'elaborazione del linguaggio naturale con sistemi automatizzati, in ogni caso, presenta delle peculiarità specifiche, che, come sarà analizzato più nel dettaglio a breve, possono spesso portare a risultati non accurati nell'attività di moderazione. Il valore di un'espressione, infatti, dipende fortemente dal contesto, dal pubblico preso a riferimento, da elementi impliciti e non è mai, in ogni caso, totalmente libero da ambiguità semantica. Sviluppare sistemi per il riconoscimento del linguaggio naturale per le lingue che non siano tra le poche più usate del mondo può non essere economicamente conveniente, né particolarmente agevole per la scarsità di materiali a disposizione con cui "allenare" il sistema. Pongono problemi specifici, inoltre, l'utilizzo di espressioni dialettali, o semplicemente scorrette dal punto di vista sintattico o grammaticale, o i contenuti ibridi eventualmente elaborati da persone viventi in un contesto di plurilinguismo[374].

3. Le criticità connesse all'automazione della *content moderation*, il ruolo dell'essere umano e le soluzioni, solo parziali, elaborate dalle piattaforme

L'automazione dell'attività di *content moderation* con gli strumenti visti al paragrafo precedenteha, come già accennato, effetti sulla possibilità di esercizio in concreto della libertà di manifestazione del pensiero che meritano un'analisi approfondita. Questo in ragione della posizione di preminenza assunta, come già detto, dagli operatori del web 2.0 nel mercato dell'informazione e della comunicazione, che ha portato le piattaforme ad assumere un ruolo essenziale per l'esercizio della libertà d'espressione, come riconosciuto da più parti in letteratura e, come si dirà, anche da alcuni provvedimenti giurisdizionali, pur con alcuni distinguo[375]. I maggiori social media, infatti,

[373] Per alcuni esempi, cfr. M. J. GARBADE, *A simple introduction to Natural Language Processing*, in *Becoming humans: Artificial Intelligence magazine*, 15 ottobre 2018, https://bit.ly/3ecJYgK (4 maggio 2022); T. DAVIDSON, D. WARMSLEY, M. MACY, I. WEBER, *Automated hate speech detection and the problem of offensive language*, 2017, http://arxiv.org/abs/1703.04009 (4 maggio 2022); P. BURNAP, M.L. WILLIAMS, *Cyber hate speech on Twitter: an application of machine classification and statistical modeling for policy and decision making*, in *Policy & Internet*, 2015, 7, 2, p. 223–242; Y. WILKS, M. STEVENSON, *The grammar of sense: using part-of-speech tags as a first step in semantic disambiguation*, in *Natural Language Engineering*, 1998, 4, 2, p. 135–143; GOOGLE CODE, *word2vec*, 2013, https://code.google.com/archive/p/word2vec/ (4 maggio 2022).

[374] Cfr., tra i molti, S. QUINTARELLI, *Content moderationcit.*,p. 119 ss; N. DUARTE, E. LLANSO, A. LOUP, *Mixed messages? The limits of automated social media content analysis cit.*, p. 12 ss; R. GORWA, R. BINNS, C. KATZENBACH, *Algorithmic content moderation cit.*, p. 4 ss.; E. LLLANSÒ, J. VON HOBOKEN, P. LEERSSEN, J. HARAMBAM, *Artificial intelligence, content moderation and freedom of expression cit.*, p. 5 ss.; W. KNIGHT, *AI's Language Problem*, in *MIT Tech. Review*, 2016, https://bit.ly/3Q2It8L (4 maggio 2022).

[375] Tra I contributi in materia si indicano, senz'animo di completezza: D. KELLER, *Making Google the censor*, in *The New York Times*, 12 giugno 2017, https://nyti.ms/35JLPqv (7 maggio 2021); K. KLONICK, *The new governors: the people, rules, and process governing online speech*, in *Harvard Law Review*, 131, 6, 2018, p. 1639 ss.; G. PITRUZZELLA, O. POLLICINO, S. QUINTARELLI, *Parole e potere cit.*; C. SUNSTEIN, *#Republic.com: divided democracy in the age of social media cit.*; J. M. BALKIN, *Free speech in the algorithmic society cit.*; E. PARISIER, *The filter bubble cit.*; E. LLLANSÒ, J. VON HOBOKEN, P. LEERSSEN, J. HARAMBAM, *Artificial intelligence, content moderation and freedom of*

connettono oggi centinaia di milioni di utenti[376], rappresentando di certo la via di comunicazione più potente e immediata per la comune persona fisica. Il numero limitato delle piattaforme, che operano in condizioni di sostanziale oligopolio, inoltre, fa sì che l'esclusione di un utente o la cancellazione di un contenuto anche da solo una di queste abbia conseguenze che non vengono totalmente mitigate dall'eventuale permanenza sulle altre. L'attività di *contentmoderation*, dunque, non si può considerare di natura meramente tecnica e interna alle piattaforme.

L'utilizzo di sistemi automatizzati per il filtro dei contenuti presenta criticità non trascurabili. Tra questedeve segnalarsi, in primo luogo, il rischio sempre presente di falsi positivi o falsi negativi nei risultati. Nonostante la rapidità del progresso tecnologico e gli importanti traguardi già raggiunti, l'accuratezza dei sistemi in esame non può mai dirsi perfetta[377]. Sono evidenti le conseguenze di natura censoria, senza una valida ragione alla base, che possono derivare dall'eccessiva moderazione di contenuti, così come i pericoli connessi a una moderazione carente.

In secondo luogo, come già accennato, il contesto in cui il sistema opera ha un ruolo preponderante, e le possibilità di errore appena menzionate crescono proporzionalmente alla sua ambiguità. Sviluppare algoritmi in grado di riconoscere notizie false o discorsi d'odio, ad esempio, obbliga a confrontarsi con le più sottili differenze terminologiche, contenutistiche e di intonazione. In più, vi sono due ulteriori elementi da tenere in considerazione. Innanzitutto, può risultare problematico definire gli stessi confini della categoria di contenuti vietata (ad esempio, distinguere tra discorsi d'odio e posizioni politiche minoritarie, ma legittime, può non essere agevole). Inoltre, può essere difficile reperire dati di qualità con cui "allenare" un sistema (si pensi al problema, già accennato,

---

expression cit.; C. CASONATO, *Intelligenzaartificiale e dirittocostituzionale: prime considerazioni cit.* e *Costituzione e intelligenza artificiale cit.*; M. FASAN, *Intelligenzaartificiale e pluralismo cit.*; M. MONTI, *Privatizzazionedellacensura e internet platforms: la libertà di espressione e i nuovicensoridell'agoràdigitale*, in *Rivistaitaliana di informatica e diritto*, 1, 2019, p. 35 ss e *La disinformazione online cit.*; M. KENNEY, J. ZYSMAN, *The rise of platform economy cit.*; J. E. COHEN, *Law for the platform economy cit.*; D.S. EVANS, R.L. SCHMALENSEE, *Matchmakers: the new economy of multisided platforms cit.*; A. GAWER, *Platforms, markets and innovation cit.*; S. , *The age of surveillance capitalism cit*. Invece, i provvedimenti giudiziari cui si fa riferimento sono, in particolare: Tribunale di Roma – sez. imprese, ord. 12 dicembre 2019, in *Diritto di internet*, con commento di A. VENANZONI, *Pluralismo politico e valore di spazio di dibattito pubblico della piattaforma social Facebook: la vicenda CasaPound*, 12 dicembre 2019, https://bit.ly/3yiXdac (7 maggio 2022); Tribunale di Roma – sez. diritti della persona e immigrazione, ord. 23 febbraio 2020, in *Questione Giustizia*, 24 febbraio 2020, https://bit.ly/3rKjyLc (7 maggio 2022); District Court S.D., New York, *Knight First Amendment Institute v. Trump, No. 1:17-cv-05205 – Order on motion for SummaryJudgment*, 23 maggio 2018, in *CourtListener.com*, https://bit.ly/3fmoaII (7 maggio 2022); U.S. Court of Appeals 2nd Circuit, *Knight First Amendment Institute v. Trump, No. 18-1691-cv*, 9 luglio 2019, in *Justia US Law*, https://law.justia.com/cases/federal/appellate-courts/ca2/18-1691/18-1691-2019-07-09.html(7 maggio 2022).

[376] Ad esempio, al gennaio del 2022 *Facebook* sfiorava i 3 miliardi di utenti, *Instagram* il miliardo e mezzo, *WeChat*il miliardo, cfr. *Mostpopular social networks worldwideas of January 2022, ranked by number of monthlyactiveusers*, Statista.com, https://bit.ly/3GXtGb0 (7 maggio 2022).

[377] S. QUINTARELLI, *Content moderation: i rimedi tecnici cit., p. 119-120;* G. CHASTEL*, Whyis Natural Language Processing still so innatural?*, in *NewtonX-accessknowledge*, 27 marzo 2018, https://bit.ly/36yUyMA (7 maggio 2022); M. J. GARBADE, *A simpleintroduction to Natural Language Processing*, in *Becominghumans: Artificial Intelligence magazine*, 15 ottobre 2018, https://bit.ly/3ecJYgK (7 maggio 2022).

del riconoscimento del linguaggio naturale in lingue minori, o in campi di applicazione estremamente specialistici e circoscritti) [378].

Infine, è stato evidenziato come i limiti delle tecnologie di moderazione online più diffuse possano avere effetti negativi sproporzionati su gruppi sociali marginalizzati, esacerbando discriminazioni già esistenti[379]. La minor efficacia di tali sistemi di fronte a contenuti espressi in versioni non standard della lingua di riferimento, ad esempio, può portare a risultati meno accurati che colpiscono determinate fasce della popolazione. È stato dimostrato come la capacità di elaborare la variante del c.d. *African American Vernacular English* sia, in molti sistemi destinati a moderare contenuti in lingua inglese, estremamente bassa [380]. La difficoltà ad operare in contesti particolarmente ambigui, inoltre, può portare a moderare contenuti che, pur contenendo termini considerabili *hatespeech*, provengano in realtà da membri delle comunità prese a bersaglio da tali insulti, che se ne appropriano nel corso della loro militanza. Il risultato finale potebbe essere, quindi, l'oscuramento di contenuti aventi finalità opposta a quella da censurare, come avvenuto in più occasioni, spesso riguardo a contenuti provenienti da attivisti del movimento LGBTQI+[381].

Di fronte a questi limiti di varia natura, è opinione condivisa che la *contentmoderation*non debba svolgersi unicamente attraverso sistemi automatizzati [382]. Sembrano condividere questa impostazione anche le piattaforme, che hanno sviluppato varie forme di intervento umano nelle decisioni attinenti all'oscuramento, alla penalizzazione nel posizionamento o al *flagging* dei contenuti in esse diffusi. Facebook, per limitarsi a un solo esempio, ha sviluppato una procedura che prevede vari livelli di possibile controllo umano sulle decisioni di moderazione, cui ora fa da supervisore ultimo l'*Oversight Board*, un organo di recente creazione, formato da esperti d'alto profilo esterni all'azienda, destinato a decidere sulle vicende particolarmente delicate e

[378] V. RUBIN, Y. CHEN, N. CONROY, *Deception detection for news, three kinds of fakes*, in *Proceedings of the American Society for Information Science and Technology*, 52, 2015; SWISS COGNITIVE – THE GLOBAL AI HUB, *Deep learning won't detect fake news, but it will give fact-checkers a boost*, 29 febbraio 2020, https://bit.ly/2yu7Q0k (7 maggio 2022); N. DUARTE, E. LLANSO, A. LOUP, *Mixed messages? cit.*, p. 16 ss.; R. GORWA, R. BINNS, C. KATZENBACH, *Algorithmic content moderation cit.*, p. 7 ss..; E. LLANSÒ, J. VON HOBOKEN, P. LEERSSEN, J. HARAMBAM, *Artificial intelligence, content moderation and freedom of expression cit.*, p. 5 ss.

[379]Cfr. ancora, ad esempio, N. DUARTE, E. LLANSO, A. LOUP, *Mixed messages? cit.*

[380]S. L. BLODGETT, B. O'CONNOR, *Racial disparity in Natural Language Processing: a case-study of social media African-American English*, Proceedings of the Fairness, Accountability,and Transparency in Machine Learning Conference, 2017, https://arxiv.org/pdf/1707.00061.pdf.

[381]Si vedano, ad esempio, D. LUX, *Facebook's hate speech policies censor marginalized users*, Wired, 14 agosto 2017; E. ROSENBERG, *Facebook blocked many gay-themed ads as part of its new advertising policy, angering LGBT groups*, The Washington Post, 3 ottobre 2018; S. GOLDING-YOUNG, *Facebook's discrimination against the LGBT community*, ACLU, 24 settembre 2020. Per le vicende di alcuni termini originariamente denigratori verso la comunità LGBTQI+, poi sempre più spesso rivendicati dai membri di quest'ultima, tra i molti cfr. E. NOSSEM, *Queer, frocia, femminiellə, ricchione et al. Localising "queer" in the Italian context*, in *Gender/Sexuality/Italy*, 6, 2019, https://doi.org/10.15781/31yc-ys20.

[382]Cfr. ad esempio S. QUINTARELLI, *Content moderation: i rimeditecnici cit.*, p. 147 ss; E. LLANSÒ, J. VON HOBOKEN, P. LEERSSEN, J. HARAMBAM, *Artificial intelligence, content moderation and freedom of expression cit.*, p. 22; DUARTE, E. LLANSO, A. LOUP, *Mixed messages? cit.*, p. 20 ss.

controverse[383]. Tali tentativi di mantenere nelle mani degli esseri umani le decisioni sulla moderazione dei contenuti non paiono, però, superare totalmente tre principali obiezioni.

In primo luogo, la quantità dei contenuti diffusi sulle piattaforme e la velocità con cui ciò avviene rendono necessario l'utilizzo massiccio di sistemi automatizzati, e di conseguenza dell'intelligenza artificiale.Non è ipotizzabile che l'attività di moderazione sia svolta solo da esseri umani, né che questi né svolgano la maggior parte, che non può che essere affidata ad algoritmi, senza alternative credibili. L'intervento umano, nei fatti, avviene in seconda battuta, come attività di ratifica e revisione della moderazione svolta dai sistemi automatizzati, e in prima battuta su segnalazione degli utenti delle piattaforme, quando trovino inopportuno un contenuto non filtrato da tali sistemi.

In secondo luogo, la qualità della moderazione svolta da esseri umani suscita diversi sospetti. Come detto, operatori in carne e ossa intervengono al momento di revisionare le decisioni dei sistemi di moderazione automatizzata; un'altra attività che li vede spesso coinvolti è la catalogazione di contenuti destinati a formare i *dataset* con cui allenare tali algoritmi. Si tratta di lavori spesso svolti in condizioni di estremo precariato da lavoratori locati in paesi del secondo e terzo mondo (i c.d. *turkers*) pagati pochi centesimi per ogni contenuto moderato. Sono sorte preoccupazioni, inoltre, per gli effetti a lungo termine delle mansioni in cui sono impiegati, consistenti, all'atto pratico, nella visione prolungata di contenuti sgradevoli e shockanti[384]. L'indignazione provocata dalla presa di coscienza, almeno parziale, delle condizioni di tali lavoratori da parte delle opinioni pubbliche dei paesi occidentali ha portato negli anni più recenti ad alcuni miglioramenti, e diverse piattaforme, oggi, affermano che la maggior parte dei loro moderatori è assunta con regolari contratti di lavoro subordinato[385]. Ciò nonostante, le criticità non sembrano del tutto superate, e la qualità dell'attività di moderazione pare comunque messa a rischio dall'enorme volume di lavoro cui i moderatori

[383]Per l'identità dei suoi membri e altre informazioni, cfr. il sito internet: https://oversightboard.com/ (9 maggio 2022). V. anche D. CASATI, M. PENNISI, *La corte suprema di Facebook: chi sono le 20 personalità che hanno deciso sul bando di Trump*, Corriere della Sera, 29 maggio 2021; A. ROBERTSON, *Go readabouthowFacebook's pseudo-Supreme Court come together*, The Verge, 12 febbraio 2021,https://bit.ly/3MD89FK (9 giugno 2022).

[384] S.T.ROBERTS, *Behind the screen: the hidden digital labor of commercial content moderation*, 2014, http://hdl.handle.net/2142/50401 (10 maggio 2022); S. QUINTARELLI, *Content moderation: irimeditecnicicit.*;A. HERN, *Revealed: catastrophic effects of working as a Facebook moderator*,The Guardian, 17 settembre 2019, https://bit.ly/2WeflS1 (10 maggio 2022); F. C. MACKENZIE, *Fear the Reaper: how content moderation rules are enforced on social media*, in *International Review of Law, Computers & Technology*, 34, 2, 2020, 128 ss.

[385]Cfr. ad esempioV. GOEL,*Facebook scrambles to police content amid rapid growth*, The New York Times, 3 maggio 2017; J. SHIEBER, *After criticism over moderation treatment, Facebook raises wages and boosts support for contractors*, TechCrunch+, 13 maggio 2019, https://tcrn.ch/3zoVIuu (10 maggio 2022). Molti dei principali *social media*, inoltre, hanno dichiarato di sposare i Principi di Santa Clara sull'eticità della moderazione dei contenuti, proposti nel 2018 da un gruppo di lavoro formato da accademici e attivisti dei diritti umani, https://santaclaraprinciples.org/ (10 maggio 2022).

umani sono chiamati e dalla velocità con cui devono prendere decisioni talvolta estremamente complesse[386].

Infine, la natura automatizzata, umana o composita delle decisioni riguardanti il filtro di contenuti lascia sullo sfondo un problema più ampio e complesso, su cui di recente si è concentrata la riflessione scientifica e che è stato oggetto di indagine da parte delle corti di alcuni paesi, Italia compresa: la riduzione in mani private di un'attività dal chiaro tenore censorio come la *content moderation*[387]. A prescindere dalle tecnologie adottate, infatti, la moderazione dei contenuti sottende scelte che implicano un controllo sul discorso pubblico, alla luce della mezionata posizione di preminenza in quest'ultimo raggiunta dalle piattaforme. Scelte che, fino al passato recente, erano considerate un'esclusiva dei poteri pubblici e a cui, negli ordinamenti democratici basati sul pluralismo ideologico, era tradizionalmente riservato un ruolo di *extrema ratio*, sottoposto a precise garanzie procedurali, e spesso a riserva di legge e di giurisdizione, qualora riguardassero i media tradizionali. Parallelamente, risultava immaginabile solo con molta difficoltà la censura della possibilità di esprimersi di un singolo individuo, privo di un ruolo definito nel mercato dell'informazione. Dalle dichiarazioni del singolo potevano, semmai, derivare conseguenze negative *ex post*, sul piano risarcitorio (per la lesione del diritto all'onore, alla reputazione, o, come visto, alle varie dimensioni dell'identità personale) o individuate da precise disposizioni penali (è il caso, ad esempio, del reato di diffamazione o della fattispecie, oggi depenalizzata, di ingiuria). Oggi, invece, il quadro appare mutato a causa del potere esercitato dalle potenti società private che gestiscono le piattaforme.

[386] A. SATARIANO, M. ISAAC, *The silent partner cleaning up Facebook for $500 million a year*, The New York Times, 31 agosto 2021; C. CRIDDLE, *Facebook moderator: every day was a nightmare*, BBC News, 12 maggio 2021, https://www.bbc.com/news/technology-57088382 (10 maggio 2022).

[387] Sul tema si rimanda, in generale, ai già citati D. KELLER, *Making Google the censorcit.*;K. KLONICK, *The new governors cit.*; G. PITRUZZELLA, O. POLLICINO, S. QUINTARELLI, *Parole e potere cit.*; C. SUNSTEIN, *#Republic.com: divideddemocracy in the age of social media cit.*; J. M. BALKIN, *Free speech in the algorithmic society cit.*; E. PARISIER, *The filterbubble cit.*; E. LLLANSÒ, J. VON HOBOKEN, P. LEERSSEN, J. HARAMBAM, *Artificial intelligence, contentmoderation and freedom of expression cit.*; C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni cit.* e *Costituzione e intelligenza artificiale cit.*; M. FASAN, *Intelligenza artificiale e pluralismo cit.*;C. M. REALE, M. TOMASI, *Libertà d'espressione, nuovi media e intelligenza artificiale cit.*; M. MONTI, *Privatizzazione della censura e internet platforms: la libertà di espressione e i nuovi censori dell'agorà digitale*, in *Rivista italiana di informatica e diritto*, 1, 2019, p. 35 ss e *La disinformazione online cit*. I provvedimenti giudiziari cui si fa riferimento sono, innanzitutto, i già citati Tribunale di Roma – sez. imprese, ord. 12 dicembre 2019; Tribunale di Roma – sez. diritti della persona e immigrazione, ord. 23 febbraio 2020; District Court S.D., New York, *Knight First AmendmentInstitute v. Trump, No. 1:17-cv-05205 – Order on motion for SummaryJudgment*, 23 maggio 2018; U.S. Court of Appeals 2nd Circuit, *Knight First AmendmentInstitute v. Trump, No. 18-1691-cv*, 9 luglio 2019.

4. La privatizzazione della censura e le sue conseguenze sul piano del diritto: il quadro normativo, il ruolo di *soft-law* e autoregolazione, le prime soluzioni giurisprudenziali

4.1 Il principio dell'intermediary liability exemption e le principali regolazioni esistenti, in materia di contentmoderation, nei vari ordinamenti

In seno al fenomeno di privatizzazione della censura brevemente descritto al paragrafo precedente si distinguono l'attività di moderazione portata avanti su delega dei poteri pubblici o in adempimento di obblighi di legge e quella messa in atto spontaneamente dalle piattaforme, al di fuori di un preciso quadro legale[388]. Deve segnalarsi che gli ordinamenti di Europa e Stati Uniti – due dei mercati in cui sono sorte e operano le principali piattaforme, e in cui la libertà d'espressione è considerata una garanzia fondamentale –fin dai primi anni dello sviluppo su larga scala della rete, hanno optato per non onerare gli operatori della responsabilità per i contenuti diffusi dagli utenti. Sia la Sec. 230 del *Communication Decency Act* americano del 1996[389] che gli artt. da 12 a 15 della Direttiva UE 2000/31/CE (c.d. *direttiva e-commerce*)[390]escludono che sulle piattaforme gravi una responsabilità generalizzata sui contenuti che veicolano, escludendo che la loro posizione possa equipararsi a quella di un editore.L'obbligo di rimuovere i contenuti in tempi rapidi, ed eventuali responsabilità in caso di inerzia, possono derivare dal venire a conoscenza, in qualunque modo, dell'illiceità di un contenuto determinato. Si tratta del principio, comune ai diversi ordinamenti, della c.d. *liability exemption* per gli intermediari digitali[391].

Le ragioni che hanno portato alla creazione della *liability exemption* stanno essenzialmente nel timore che l'individuazione di responsabilità specifiche avrebbe portato alla moderazione di contenuti in misura maggiore al necessario e a forme di autocensura da parte degli utenti per la paura di incorrere nelle sanzioni della piattaforma. L'esenzione da responsabilità delle piattaforme

[388] Sulle questioni sollevate dalla crescente priatizzazione dell'attività censoria, cfr. D. KELLER, *Making Google the censor cit.*; M. MONTI, *Privatizzazione della censura e internet platforms: la libertà di espressione e i nuovi censori dell'agorà digitale*, p. 36 ss.; *La disinformazione online cit*., p. 295 ss.

[389]U.S. Communication Decency Act of 1996 – Title V of U.S. Telecommunication Act of 1996.La Sec. 230 recita: «No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider».Per un commento si veda J. KOSSEFF, *The twenty-six words that created the internet cit.*

[390]Direttiva 2000/31/CE del Parlamento europeo e del Consiglio dell'8 giugno 2000, *relativa a taluni aspetti giuridici dei servizi della società dell'informazione, in particolare il commercio elettronico, nel mercato interno*.

[391] In letteratura, tra i molti, si indicano J. RIORDAN, *The liability of internet intermediaries*, Oxford, 2016; A. KUCZERAWY, *Intermediary liability & freedom of expression: recent developments in the EU notice & action*, in *Computer Law & Security Review*, 31, 1, 2015, p. 46-56; R. J. MANN, S. R. BELZLEY, *The promise of internet intermediaryliability*, in *William and Mary Law Review*, 47, 1, 2005, p. 239-308; P. VAN EECKE, *Online Service Providers and Liability: A Plea for a Balanced Approach*, in *Common Market Law Review*, 48,2011, p. 1455 e ss; O. POLLICINO, *Tutela del pluralismo nell'era digitale: ruolo e responsabilità degli Internet service provider*, in *Percorsi costituzionali*, 2014, 1, p. 45-74; P. SANNA, *Il regime di responsabilità dei providers intermediari di servizi della società di informazione*, in *Responsabilità civile e previdenza*, 1, 2004, p. 279-302; A. HEUNG, R.H. WEBER, *Internet governance and the responsibility of Internet Service Providers*, in *Wisconsin International Law Journal*, 26, 2, 2008, p. 403-477; G. F. FROSIO, *Reforming intermediary liability in the platform economy: a European digital single market strategy*, in *Northwestern University Law Review*, 112, 2017, p. 19 ss.

è, peraltro, messa in crescente discussione a causa del mutamento che ha attraversato il contesto in cui è nata, che vede le piattaforme svolgere attività di indicizzazione dei contenuti e profilazione degli utenti sempre più raffinate e ottenere ricavi pubblicitari sempre più ampi, facendo dubitare di un sistema che massimizza i profitti che esse ottengono dai materiali diffusi dagli utenti, esonerandole al contempo da obblighi [392]. Il presentarsi di questioni sempre più urgenti relativamente alla moderazione dei contenuti e i ripetuti episodi in cui materiali diffusi nelle reti sociali sono sembrati avere un ruolo fondamentale nell'esarcerbarsi di tensioni sociali non hanno fatto che accrescere le perplessità in materia[393]. Devono segnalarsi interessanti ipotesi di riforma, una di particolare rilievo a livello europeo, nel quadro della proposta, risalente al 2020, di un *Digital Service Act*[394], che sarà analizzata più avanti.

Posta la mancanza di un obbligo generalizzato di controllo, la moderazione di contenuti in adempimento di obblighi di legge avviene in ipotesi puntuali, previste da normative specifiche di alcuni ordinamenti nazionali. Così, ad esempio, nel sistema italiano è prevista la compilazione, da parte di distinte autorità pubbliche, di *blacklist* di siti web in cui si diffondano contenuti illeciti (si tratta, ad esempio, di siti pedopornografici o che diffondono contenuti attinenti ad attività terroristiche) che vengono trasmesse agli *internet service provider* per il loro oscuramento[395]. La Legge 71/2017, in materia di c.d. cyberbullismo, invece, delega esplicitamente alle piattaforme valutare se un contenuto rientri o meno nella categoria, e in caso positivo ne impone la rimozione

[392]M. D. SMITH, M. VAN ALSTYNE, *It's time to update Section 230*, in *Harvard Business Review*, 12 agosto 2021, https://hbr.org/2021/08/its-time-to-update-section-230 (10 maggio 2022); J. ADETUNJI, *Tech giants need to take more responsibility for the advertising that makes them billion*, The Conversation, 7 dicembre 2018, M. REARDON, *Section 230: how it shields Facebook and why Congress wants changes*, CNET, 6 ottobre 2021, https://cnet.co/3HaxndH (10 maggio 2022). Inoltre si vedano, in chiave in parte critica, le osservazioni di G. F. FROSIO, *Whykeep a dog and barkyourself? From intermediary liability to responsibility*, in *International Journal of Law and Information Technology*, 26, 2018, p. 1-33; S. STALLA-BOURDILLON, *Liability exemptions wanted! Internet intermediaries' liability under Uk law*, in *Journal of International Commercial Law and Technology*, 7, 4, 2012.

[393]In particolare, nel mondo occidentale hanno suscitato scalpore le conseguenze della diffusione sui *social network* di disinformazione riguardo alla pandemia di Covid19 e il ruolo giocato dalle piattaforme nell'assalto al Campidoglio degli Stati Uniti d'America del 6 gennaio 2021, cfr. G. CADALANU, *Coronavirus, le bufale sull'esercitazione Defender Europe e l'invio di soldati USA*, la Repubblica, 13 marzo 2020; S. LEE MYERS, *China spins tale that the U.S. armystarted the coronavirus epidemic*, TheNew York Times, 13 marzo 2020; J. TEMPERTON, *How the 5G coronavirus conspiracytheorytorethrough the internet*, Wired, 6 aprile 2020; EUVSDISINFO, *Report- Repeating a liedoesnotmakeittrue*, 9 aprile 2020, https://bit.ly/3cgfie2 (10 maggio 2022); *U.S. Capitolriot*,New York Times (online), https://nyti.ms/2TL1iEY (20 luglio 2021); *41 minutes of fear: a video timeline from inside the Capitolsiege*,The Washington Post (online), https://wapo.st/3C2P3oA (10 maggio 2022).

[394] COMMISSIONE EUROPEA, *Proposta di Regolamento al Parlamento Europeo e al Consiglio relativo a un mercato unico dei servizi digitali (legge sui servizi digitali) e che modifica la direttiva 2000/31/CE*, 15 dicembre 2020, COM(2020) 825 final.

[395] La Legge n. 38 del 6 febbraio 2006, *Disposizioni in materia di lotta contro lo sfruttamento sessuale dei bambini e la pedopornografia anche a mezzo Internet*, ad esempio, ha istituito il Centro nazionale per il contrasto alla pedopornografia sulla rete INTERNET, con cui gli ISP collaborano per l'oscuramento delle pagine che diffondano materiale illecito; la Legge n. 43 del 2015, invece, ha previsto la compilazione, da parte dell'organo del Ministero dell'interno per la sicurezza e la regolarità dei servizi di telecomunicazione, di una lista di siti utilizzati per il compimento dei reati di cui agli arrt. 270-bis e 270-sexies c.p.

entro 48 ore[396].Nel resto d'Europa spicca il caso della Germania, che, nel 2017, ha emanato una normativa (la *Netzwerkdurchsetzungsgesetz*, traducibile come *Legge sull'applicazione della rete*) che impone obblighi specifici ai principali servizi di social network, il principale dei quali consiste nella rimozione, entro 24 ore dal ricevimento di una notifica in proposito, dei contenuti "manifestamente illegali"[397]. Tale periodo si estende a sette giorni qualora la decisione richieda una valutazione più complessa, ed è prevista la possibilità, per le piattaforme, di incaricare di tali valutazioni un organo ad hoc finanziato unicamente con loro risorse, che deve ricevere la preventiva approvazione da parte delle autorità pubbliche. A seguito di una modifica risalente al 2021, la legge obbliga le piattaforme alla denuncia alla polizia federale di determinati crimini d'odio di cui vengano a conoscenza nell'attività di moderazione. Una normativa simile, proposta in Francia nel 2020, la *loidu 24 juin 2020 visant à luttercontrelescontenushaineuxsur internet* (più nota come *loiAvia*, dal cognome della prima firmataria) [398] è stata, invece, in larga parte censurata dal ConseilConstitutionnel, che ha considerato restrizioni eccessive della libertà d'espressione l'imposizione di termini temporali stringenti alle piattaforme per la rimozione dei contenuti (si trattava, anche in tal caso, di 24 ore)e il mancato coinvolgimento dell'autorità giurisdizionale nell'intero procedimento censorio disegnato dalla legge[399]. Alle legislazioni di questo tipo si aggiunge, sullo scenario europeo, la procedura di deindicizzazione di contenuti dai motori di ricerca prevista dalla sentenza *Google Spain* a tutela del diritto all'oblio, vista nei paragrafi precedenti.

Al contrario che in Europa, negli Stati Uniti non si rinvengono normative volte ad imporre obblighi di questo genere alle piattaforme. Regolamentazioni che impongono agli operatori di internet condotte censorie – o che impediscono radicalmente di operare a talune di queste – sono invece presenti in una pluralità di ordinamenti estranei alla tradizione occidentale[400]. Si tratta, nella quasi

[396] Per la precisione, l'art. 2 della Legge n. 7 del 29 maggio 2017, *Disposizioni a tutela dei minori per la prevenzione ed il contrasto del fenomeno del cyberbullismo*, attribuisce al minore ultraquattordicenne o all'esercente la responsabilità genitoriale la facoltà di richiedere al gestore di un sito internet o social media l'oscuramento, entro 48 ore, di un contenuto che ritiene rientrare nella categoria del cyberbullismo.

[397] Cfr. *Netzwerkdurchsetzungsgesetz*, Beschlussempfehlung und Bericht, Deutscher Bundestag: Drucksache [BT] 18/13013, 28 juni 2017, § 3(2). Per un resoconto in lingua inglese dei contenuti della legge, si veda D. LEE, *Germany's NetzDG and the Threat to Online Free Speech*, in *Yale Law School – MFIA*, 10 ottobre 2017, https://bit.ly/3TQnyXX (30 ottobre 2021).

[398] LOI n. 2020-766 du 24 juin 2020 *visant à lutter contre les contenus haineux sur internet*, Journal officiel "Lois et Décrets" n. 0156 du 25 juin 2020.

[399] Conseil Constitutionnel, Décision n° 2020-801 DC du 18 juin 2020. In letteraturacfr. C. SICCARDI, *La "loi Avia". La leggefrancesecontrol'odio online (o quelloche ne rimane)*, in M. D'AMICO, C. SICCARDI, *La Costituzione non odia: conoscere, prevenire, contrastarel'hate speech online*, Torino, 2021, p. 167-183; D. STEIGER, *Protecting Democratic Elections Against Online Influence via "Fake News" and Hate Speech – The French Loi Avia and Loi No. 2018–1202, the German Network Enforcement Act and the EU's Digital Services Act in Light of the Right to Freedom of Expression*, in M. KOTZUR, S. SCHIEDERMAIR, D. STEIGER, M. WENDEL (A CURA DI),*Theory and practice of the European Convention on Human Rights*, 2021, p. 165-215.

[400] La Repubblica Popolare Cinese, ad esempio, esercita un controllo penetrante sulla rete internet, impedendo radicalmente l'accesso ad alcune delle principali piattaforme occidentali, tra cui Facebook, Twitter o Youtube; misure analoghe sono in vigore in Iran, Arabia Saudita, Corea del Nord e diversi altri paesi non democratici. In alcuni ordinamenti norme formalmente concepite per combattere la disinformazione o certe forme di propaganda (ad esempio

totalità dei casi, di sistemi apertamente autocratici o comunque non totalmente rispettosi degli standard democratici, in cui le inedite possibilità di espressione e autoorganizzazione dei cittadini comuni dischiuse dall'internet 2.0 incontrano l'ostilità del potere costituito. Ciò esclude che possano essere di particolare interesse ai fini di questo lavoro, posto che le soluzioni da essi elaborate sono sintomi di una cultura giuridica in cui la libera manifestazione del pensiero non è considerata un diritto fondamentale, bilanciabile solo con interessi di pari rango, ma una componente della vita pubblica sacrificabile per una pluralità di esigenze, in primo luogo il mantenimento di un determinato *status quo*.

4.2 I principali strumenti di soft-law in materia di content moderation sullo scenario europeo, l'evoluzione delle regole interne delle piattaforme e le sue ragioni: il caso delle notizie false

Accanto a queste ipotesi puntuali, nelle quali le piattaforme adempiono a obblighi imposti dalla legge, vi è la gran parte dell'attuale attività di *content moderation*, messa in atto al di fuori di un preciso quadro legale. Talvolta, essa avviene sulla base di documenti di autoregolazione elaborati e sottoscritti dalle piattaforme stesse, con il coordinamento, per quanto riguarda lo scenario europeo, della Commissione. In tali casi, dunque, ferma l'autonomia con cui operano le piattaforme, è rinvenibile quantomeno un *endorsment* da parte dei poteri pubblici. Ad ogni modo, deve evidenziarsi che si tratta di documenti contenenti dichiarazioni d'intenti molto generali, che ciascuna piattaforma sottoscrive volontariamente, non idonei a costituire obblighi legalmente vincolanti e che non prevedono procedure d'*enforcement* in caso di mancato rispetto degli impegni presi. Si tratta, essenzialmente, del *Codice di condotta per contrastare l'illecito incitamento all'odio online*, risalente al 2016, cui oggi aderiscono 9 piattaforme[401], e del *Code of practice against disinformation*, siglato nel 2018 e attualmente condiviso da 6 operatori[402].

---

volta a incitare a condotte presuntamente "terroristiche") hanno, sul piano sostanziale, l'effetto di sottoporre a forme di censura significative i contenuti che circolano online (è il caso, ad esempio, di Russia e Turchia). Informazioni approfondite sullo stato della censura di internet nei diversi paesi del mondo sono reperibili, tra i molti, in REPORTERS WITHOUT BORDERS, *2022 World Press Freedom Index*, https://rsf.org/en (10 maggio 2022); P. BISCHOFF, *Internet censorship 2022: a global map of internet restrictions*, Comparitech – Report, 25 gennaio 2022, https://bit.ly/3QirmzA (10 gennaio 2022).

[401] Il *Codice di condotta* è stato siglato il 31 maggio 2016 dalla Commissione con le piattaforme Youtube, Facebook, Twitter e Microsoft. In seguito, tra il 2018 e il 2019, vi hanno aderito Instagram, Google+, Dailymotion, Snap e Jeuxvideo.com. Il testo del documento è consultabile alla pagina web della Commissione Europea: https://bit.ly/3aX57iM (13 maggio 2022). Cfr. ancheEUROPEAN COMMISSION, *Progress on combating hate speech online through the EU Code of conduct*, 27 settembre 2019, https://bit.ly/3NMA4Vl (13 maggio 2022).

[402]Il *Code of practice*è una raccolta di standard in materia di contrasto alla disinformazione elaborato nel quadro della Comunicazione della Commissione Europea *Tackling online disinformation: a Europeanapproach*, 26 aprile 2018, COM/2018/236 final, https://bit.ly/3zwChQj (13 maggio 2022) siglata nell'ottobre 2018 da Facebook, Google, Twitter e Mozilla. A queste piattaforme si sono unite Microsoft, nel 2019, e TikTok, nel 2020. Il testo del documento è disponibile nel sito della Commissione: https://bit.ly/3xKTxQw (13 maggio 2022).

Al di fuori dalla vaga cornice di questi strumenti di *soft-law*, l'attività di moderazione si svolge unicamente sulla base delle condizioni d'uso e dei c.d. standard della comunità delle piattaforme, nel quadro di una relazione con gli utenti che esse concepiscono come meramente contrattuale[403]. È significativo notare come, pressoché in tutti i casi, tali regole interne non menzionino fonti normative vincolanti statali o sovranazionali, nell'evidente intento di costituire un sistema autosufficiente[404]. A partire da questa caratteristica, sono stati avanzati paragoni suggestivi, sostenendo che il controllo delle piattaforme sul comportamento degli utenti somigli sempre più a un esercizio di sovranità (tesi, questa, che ha acquisito vigore con la creazione, da parte di *Facebook*, dell'*Oversight Board* già menzionato, subito soprannominato la "corte suprema" della piattaforma)[405]. A prescindere da questa considerazione, deve evidenziarsi un dato pragmatico: i grandi operatori di internet hanno un chiaro interesse nel tentare di risolvere i problemi generati dalla moderazione dei contenuti unicamente in base alle condizioni contrattuali che loro stessi predispongono. Ciò, infatti, permette loro di conservare gli attuali spazi di autonomia, e di evitare che gli ordinamenti in cui operano li gravino di responsabilità su quanto viene diffuso dagli utenti in misura maggiore rispetto al quadro giuridico attuale. Un rischio, quest'ultimo, da essi percepito come sempre più urgente, poste le crescenti discussioni attorno al principio della *intermediary liability exemption* già viste. I *social media*, dunque, paiono animati più dall'interesse squisitamente pratico di evitare di dover sopportare nuovi vincoli imposti dai sistemi normativi tradizionali, che dalla volontà di sviluppare un proprio corpo di regole autonomo e sovrano. In sintesi, pare fondato

[403] M. MONTI, *Privatizzazione della censura e internet platforms cit.*, p. 39 ss. distingue le forme di censura analizzate finora in "censura privata *de facto*", messa in atto dalle piattaforme al di fuori da ogni quadro normativo di riferimento, "censura privata *de iure* funzionale", in adempimento di precisi obbighi di legge, "censura privata *de iure* sostanziale", realizzata nel quadro di orientamenti normativi generali che lascino alle piattaforme il giudizio sulla liceità o meno di un determinato contenuto.

[404] Si vedano, ad esempio, gli standard della comunità di Facebook: https://bit.ly/3OucigX (13 maggio 2022); Twitter: https://bit.ly/3NOviqm (13 maggio 2022); TikTok: https://bit.ly/3NHQNJn (13 maggio 2022), le uniche a fare un generico riferimento, in un paragrafo, a contenuti relativi a *illegalactivities*, senza specificare quali ordinamenti siano da prendere a riferimento per definirne i confini e, anzi, delimitandoli loro stesse con fitti elenchi di tipologie di contenuti vietate.

[405] V. ad esempio A. ROBERTSON, *Go readabouthowFacebook's pseudo-Supreme Court come together cit.*; A. ALÙ, *Oversightboard di Facebook alla prima prova: così si disvela il suo ruolo*, Agenda digitale, 1febbraio 2021; J. D'ALESSANDRO, *Nasce la "Corte suprema" di Facebook. Indipendente, giudicherà le scelte del social network*, la Repubblica, 7 maggio 2020; S. CORSI, *Nasce la Corte suprema di Facebook: si chiama Oversight Board*, in Cyberlaws, 30 novembre 2020. Riguardo alle possibli suggestione relative a una parziale "statualità" di Facebook, è interessante notare come la piattaforma, da un lato, come già visto, ignori totalmente, nei propri standard della comunità, le leggi nazionali dei paesi in cui opera, dall'altro prenda in considerazione determinati atti di diritto internazionale di hard-law e soft-law, a cui dichiara di ispirare i propri comportamenti in apposite *policy* interne. È dunque la stessa piattaforma a decidere di aderire a tali strumenti normativi, in un modo che non può non ricordare il comportamento degli stati sovrani nei confronti delle dichiarazioni internazionali dei diritti. Cfr., per un esempio, *Facebook corporate human rights policy*, 16 marzo 2021, https://about.fb.com/news/2021/03/our-commitment-to-human-rights/ (13 maggio 2022). Sul punto, inoltre, sia consentito rinviare a L. RINALDI, *Le piattaforme tra diritto pubblico e diritto privato: libertà d'espressione, discorso politico e social network in alcuni casi recenti tra Italia e Stati Uniti*, in *Gruppo di Pisa. Dibattito aperto sul Diritto e la Giustizia costituzionale*, Quad. Monografico n. 3 - fascicolo 2, 2021, p. 223 ss.

affermare che *Facebook*, più che a creare un suo tribunale, sia interessato ad evitare quelli, già esistenti, degli stati nazionali.

Volgendo lo sguardo alla condotta in concreto delle piattaforme negli ultimi anni, si trovano agevolmente indizi di quanto appena affermato. Il rigore e la severità della moderazione di determinati contenuti, infatti, sono cresciuti di pari passo all'intensità dei problemi che la loro condivisione si è dimostrata in grado di generare, e alle discussioni sul regime giuridico da applicare agli intermediari di internet che ne sono derivate. Paradigmatico, da questo punto di vista, è stato il caso della diffusione di disinformazione, un ambito particolarmente delicato, posto che, nella gran parte degli ordinamenti democratici, la semplice pubblicazione online di un'affermazione falsa non rappresenta, di per sé sola, un illecito[406]. Essa è avvenuta in modo pressoché incontrollato e nel disinteresse generale fino a quando non è risultata evidente la possibilità di condizionare, attraverso un uso strumentale dei *social network*, ampi settori dell'opinione pubblica e la stessa genuinità di elezioni e referendum (le elezioni presidenziali americane del 2016, il referendum sulla c.d.*Brexit* e le vicende della compagnia *Cambridge Analytica* appaiono, da questo punto di vista, paradigmatiche)[407]. Un ulteriore contributo alla presa di coscienza dei possibili effetti di un uso distorto delle piattaforme è venutodalloropossibileruolo in alcuni sanguinosi eventi avvenuti nello stesso periodo, in primo luogo la persecuzione del popolo Rohingya in Myanmar, secondo molti analisti favorita dalla diffusione, principalmente su *Facebook*, di disinformazione ed *hate speech*[408]. Il risultato di questa acquisizione di consapevolezza è stato un mutamento dell'approccio alla moderazione di tali contenuti a partiredalla seconda metà degli anni '10 del 2000, che ha portato alla definizione di appositi standard interni in materia di *fake news* da parte delle piattaforme e alla creazione, sullo scenario europeo, del menzionato *Code of practice against disinformation*. In tale fase, tuttavia, l'oscuramento dei contenuti considerati *fake news* è stato in generale escluso, preferendo soluzioni non apertamente censorie che limitassero la loro viralità, come penalizzazioni

[406] Nell'ordinamento italiano, ad esempio, il falso *ex se* è, in generale, penalmente neutro. Infatti, la mera dichiarazione falsa, qualora non sia volta a ledere determinati beni giuridici (come la pubblica fede nei delitti di falso ideologico, o l'ordine pubblico nella contravvenzione di cui all'art. 656 c.p.), non è sanzionata. Questione risalente e controversa è, invece, se la menzogna sia, almeno parziale, protetta dalla garanzia costituzionale della libera manifestazione del pensiero: sembra riconoscere un margine di tutela P. BARILE, *Diritti dell'uomo e libertà fondamentali*, Bologna, 1984, 229, secondo cui «neppure la diffusione di notizie false può essere considerata illecita in sé e per sé»; di opinione contraria invece C. ESPOSITO, *La libertà di manifestazione del pensiero nell'ordinamento italiano*, Milano, 1958, p. 36-37.

[407]Cfr. H. Berghel, *Malice Domestic: The Cambridge Analytica Dystopia,* Computer, 51, 5, 2018, p. 84–89; e ancoraH. MARSHALL, A. DRIESCHOVA, *Post-truth politics in the UK's Brexit referendum cit.*; R. KÜBLER, K. PAUWELS, K. MANKE, *How Social Media Drove the 2016 US Presidential Election cit.*; A. R. DOSHI, S. RAGHAVAN, R. WEISS, E. PETITT, *How the supply of fake news affected consumer behavior during the 2016 US election cit.*

[408]Cfr. ad es. C. FINK, *Dangerous speech, anti-muslim violence, and Facebook in Myanmar*, in *Journal of International Affairs*, 71, 1, 5, 2018, p. 43–52; J. WHITTEN-WOODRINGETAL., *Poison If You Don't Know How to Use It: Facebook, Democracy, and Human Rights in Myanmar*, in *The International Journal of Press/Politics*, 25, 3, 2020, p. 407–425; T. BURRETT,*Journalism in Myanmar: Freedom, Facebook and fake news*, in J. MORRISON, J. BIRKS, M. BERRY (A CURA DI),*The Routledge Companion to Political Journalism*, London-New York, 2022.

nell'indicizzazione e limitazioni alla possibilità di ricavare profitti pubblicitari da tali materiali. Da ultimo, e in particolar modo in coincidenza del proliferare di disinformazione riguardante l'epidemia di Covid-19 scoppiata sul finire del 2019, le piattaforme hanno avviato un'intensa attività di rimozione di tali materiali, una volta identificati, alla luce del pericolo per la salute pubblica ad essi potenzialmente collegato[409].

L'attuale politica di *content moderation* delle notizie false, frutto dell'evoluzione nel comportamento delle piattaforme appena descritta e dell'esperienza accumulata nel corso dell'emergenza pandemica, vede prevalere una soluzione intermedia: le regole interne della maggior parte dei *social media* non optano per l'oscuramento generalizzato delle *fake news*, al contrario che per altri tipi di contenuto, come i discorsi d'odio. Gli operatori di internet concentrano i loro sforzi sulle menzionate strategie di penalizzazione algoritmica e pubblicitaria, limitandosi ad oscurare gli account che si dimostrano superdiffusori di disinformazione, e non i singoli contenuti, spesso diffusi, peraltro, da utenti in buona fede. Allo stesso tempo, però, la rimozione è spesso prevista a tutela di primari interessi pubblici, come la salute collettiva, nel già visto caso del Covid-19, e alcuni diritti fondamentali individuali[410]. Problemi specifici, poi, pone la diffusione di c.d. *deepfake*, materiali estremamente realistici ritraenti persone in circostanze in realtà mai avvenute,

[409] L'evoluzione degli standard della comunità di Facebook offre un'ottima testimonianza di questo cambiamento. Prima dello scoppio della pandemia, fino al dicembre 2019, la principale dichiarazione reperibile in materia di *fake news* era questa, nella versione inglese, che la piattaforma stessa considera "ufficiale": «Reducing the spread of false news on Facebookis a responsibilitythatwe take seriously. We also recognise that this is a challenging and sensitive issue […]. For these reasons, we don't remove false news from Facebook, but instead significantly reduce its distribution by showing it lower in the News Feed». In un aggiornamento redatto dopo la diffusion dellamalattia e la conseguenteproliferazione di disinformazione, invece, la societàdichiarava per la prima volta l'intenzione di rimuovere i contenutiveicolantinotizie false: «We are working to remove content that has the potential to contribute toreal-world harm, including through our policies prohibiting the coordination of harm, the sale of medical masks and related goods, hate speech, bullying and harassment, and misinformation that contributes to the risk of imminent violence or physical harm». La versione attuale (giugno 2022) è notevolmente più complessa e strutturata di quelle precedenti. Contiene norme specifiche per la disinformazione riguardante le elezioni, i vaccini o, in generale, emergenze sanitarie, e dichiara, tra le altre cose, che la piattaforma «remove misinformation or unverifiable rumors that expert partners have determined are likely to directly contribute to a risk of imminent violence or physical harm to people», cfr. *Misinformation – policy detail*, https://transparency.fb.com/policies/community-standards/misinformation/ (13 maggio 2022). Gli standard della comunità in vigore a dicembre 2019 sono consultabili al link: https://transparency.fb.com/de-de/policies/community-standards/false-news/ (13 maggio 2022). Il citato aggiornamento a causa della diffusione del Covid19 al link: https://www.facebook.com/help/instagram/477434105621119 (13 maggio 2022).

[410] Oltre all'attuale versione degli standard della comunità di Facebook-Meta, citata alla nota precedente, può farsi l'esempio dell'attuale versione delle *Twitter rules*, in cui, ad esempio, la piattaforma dichiara che: «tweets that share misleading media are subject to removal under this policy ifthey are likely to cause seriousharm. Some specific harms we consider include: threats to physical safety of a person or group; incitement of abusive behavior to a person or group; risk of mass violence or widespread civil unrest; risk of impeding or complicating provision of public services, protection efforts, or emergency response; threats to the privacy or to the ability of a person or group to freely express themselves or participate in civic events» o che: «we will label or remove false or misleading information about how to participate in an election or other civic process», cfr: https://help.twitter.com/en/rules-and-policies/twitter-rules (13 maggio 2022). In modo simile, TikTok, nelle proprie *community guidelines*, dichiara: «We will remove misinformation that causes significant harm to individuals, our community, or the larger public regardless of intent. Significant harm includes serious physical injury, illness, or death; severe psychological trauma; large-scale property damage, and the undermining of public trust in civic institutions and processes such as governments, elections, and scientific bodies», cfr. https://www.tiktok.com/community-guidelines?lang=en#37 (13 maggio 2022).

potenzialmente estremamente lesivi della dignità dei soggetti coinvolti, contro i quali diverse piattaforme dichiarano di agire con particolare rigore[411].

4.3 Le principali criticità della content moderation svolta in autonomia dalle piattaforme: incoerenza, scarsa trasparenza, mancanza di proporzione, ruolo eccessivo dell'automazione

L'attività di moderazione svolta in autonomia dalle piattaforme con le modalità viste finora ha spesso suscitato discussioni e perplessità, soprattutto per l'elevata discrezionalità che esse sembrano esercitare, con risultati finali spesso incoerenti. Nonostante gran parte dei social media, come già detto, offrano importanti rassicurazioni sul coinvolgimento di operatori umani nell'attività di moderazione, sui grandi numeri le decisioni in materia di oscuramento di contenuti e limitazione o chiusura di determinati account appaiono di frequente poco trasparentie frutto della sola automazione. I loro destinatari, inoltre, spesso non sembrano dotati di procedure con cui far sentire la propria voce senza sforzi irragionevoli[412]. Ciò è vero, in particolare, in coincidenza di eventi particolari, di fronte ai quali pare chiaro, nonostante la difficoltà di reperire, sul punto, dati approfonditi, che le piattaforme aumentino i propri sforzi in materia di moderazione dei contenuti, anche al costo di diminuire la precisione di quest'ultima[413]. È innegabile che comportamenti di

[411] Le *Twitter rules* citate alla nota precedente, ad esempio, includonotrai «misleading media» passibili di rimozione: «media depicting a real person have been fabricated or simulated, especially through use of artificial intelligence algorithms». Facebook, invece, negli standard della comunità dichiara: «media can be edited in a variety of ways. In many cases, these changes are benign, such as content being cropped or shortened for artistic reasons or music being added. In other cases, the manipulation is not apparent and could mislead, particularly in the case of video content. We remove this content because it can go viral quickly and experts advise that false beliefs regarding manipulated media often cannot be corrected through further discourse. We remove videos under this policy if specific criteria are met: (1) the video has been edited or synthesized, beyond adjustments for clarity or quality, in ways that are not apparent to an average person, and would likely mislead an average person to believe a subject of the video said words that they did not say; and (2) the video is the product of artificial intelligence or machine learning, including deep learning techniques (e.g., a technical deepfake), that merges, combines, replaces, and/or superimposes content onto a video, creating a video that appears authentic». Non manca chi consideriquestetuteleinsufficienti, cfr. T. ROMM, D. HARWELL, I. STANLEY-BECKER, *Facebook bans deepfakes, but new policy could not cover controversial Pelosi video*, The Washington Post, 7 gennaio 2020.

[412] Si vedano ancora R. GORWA, R. BINNS, C. KATZENBACH, *Algorithmic content moderation cit.*; E. LLLANSÒ, J. VON HOBOKEN, P. LEERSSEN, J. HARAMBAM, *Artificial intelligence, content moderation and freedom of expression cit.*; F. C. MACKENZIE, *Fear the Reaper: how content moderation rules are enforced on social media cit*.

[413]Le statistiche più comuni in materia di attività di moderazione, infatti, raggruppano i contenuti moderati solamente in macroaree, rendendo difficile ipotizzare correlazioni tra eventi puntuali e variazioni nei flussi. Ciò nonostante, è agevole notare, ad esempio, come la moderazione dei contenuti considerati *hatespeech*da Facebook-Meta abbia subito, negli anni, variazioni ragguardevoli (con oscillazioni spesso superiori al 100% tra un trimestre e l'altro) e, in generale, un deciso aumento dei contenuti moderati negli ultimi anni: https://bit.ly/3aS55bO (dati Statista - 13 maggio 2022). Possono farsi le stesse considerazioni in materia di contenuti considerati, a vario titolo, rappresentazione o incitazione alla violenza: https://bit.ly/3NSYY5G (dati Statista - 13 maggio 2022). Così non è stato per altre tipologie di contenuti, come quelli sessualmente espliciti, riguardo ai quali l'intensità della moderazione si è mantenuta molto più costante: https://bit.ly/3MHodqg (dati Statista - 13 maggio 2022). Ciò fa desumere che, in periodi ben determinati, la moderazione dell'*hatespeech*o dei contenuti violentisi faccia all'improvviso più stringente, la quantità di tali contenuti aumenti bruscamente, o entrambe le cose. La scarsa trasparenza delle piattaforme sui risultati della moderazione dei

questo tipo abbiano già avuto un effetto non trascurabile sul *marketplace of ideas*[414] di alcune società democratiche: è paradigmatico l'esempio di quanto avvenuto nelle prime settimane del 2021, in cui è stata limitata, con particolare energia, la diffusione di contenuti che supportavano la condotta, a molti parsa apertamente eversiva, del presidente uscente degli Stati Uniti Donald Trump, sfociata nel tristemente noto assalto al Campidoglio del 6 gennaio[415]. Si tratta di misure prese per chiare finalità di ordine pubblico, che rappresenta da sempre uno dei valori posti a bilanciamento della libertà d'espressione, anche negli ordinamenti in cui la tutela del *free speech* è particolarmente forte (paradigmatico il caso americano, in cui vige la nota dottrina del *clear and present danger*)[416]. La condotta delle piattaforme potrebbe, dunque, apparire ampiamente giustificata a un'analisi sommaria. Un fatto, però, avrebbe forse meritato maggiori riflessioni: la stretta in materia di *content moderation* registrata in quei giorni ha finito per colpire, in maniera spesso disordinata, un gran numero di cittadini comuni, e non è agevole sostenere che i giganti del web, come forse avrebbero dovuto, abbiano concentrato i loro sforzi solo su pagine e account responsabili di generare e diffondere su larga scala i contenuti interessati[417]. Operatori spesso finanziati da centri d'interesse vicini al presidente Trump, o che comunque ritenevano opportuno investire risorse per sostenerlo, e la cui unica funzione sui *social* è la diffusione di tali materiali, spesso consistenti in disinformazione o discorsi d'odio. Uno schema, peraltro, che si è ripete inalterato in tutte le situazioni (elezioni e referendum, conflitti armati, sommosse popolari, ecc.) in

---

contenuti è da tempo sotto accusa, cfr. ad. es. J. C. YORK, C. MCSHERRY, *Content Moderation Is Broken. Let Us Count the Ways*, Electronic Frontier Foundation, 29 aprile 2019, https://bit.ly/2DEAafI (14 maggio 2022).

[414] La notissima espressione, patrimonio della letteratura giuridica americana, si fa comunemente risalire al *dissent* del giudice Oliver Holmes in *Abrams vs US*, 25 U.S. 616 1919.

[415] G. BENSINGER, *Now social media grows a conscience? Facebook and Twitter are taking action. It's too little, too late*, The New York Times, 13 gennaio 2021;*How big tech companies responded to the storming of the Capitol*, The New York Times, 11 gennaio 2021; L. CURINI, *Da Trump a Q-Anon, se la censura è un boomerang (anche per gli 007)*, in *Formiche.net*, 10 gennaio 2021, https://bit.ly/3zCjROn (14 maggio 2022).

[416] La prima teorizzazione della *clear and present danger doctrine* si deve, anche in questo caso, a un'*opinion* di Justice Holmes in Schenck vs US, 249 US 47 1919. In letteratura, tra i moltissimi riferimenti possibili, si rimanda, per un inquadramento dell'evoluzione della dottrina dalle origini all'età digitale, a W. MENDELSON, *Clear and Present Danger--From Schenck to Dennis*, in *Columbia Law Review* 52, 3, p. 312 ss., 1952; F. R. STRONG, *Fifty Years of 'Clear and Present Danger': From Schenck to Brandenburg - and Beyond*, in *The Supreme Court Review*, 1969 p. 41–80; O. POLLICINO, *La prospettiva costituzionale sulla libertà d'espressione nell'era di internet cit*.

[417] Non è agevole reperire dettagli sulle modalità e la scala dell'attività di moderazione messa in atto dalle piattaforme nelle settimane successive ai fatti di Capitol Hill, perché non sono mai stati diffusi dati sul punto. Il fatto che siano stati compiuti sforzi inediti, che hanno necessariamente coinvolto un gran numero di account di persone comuni, è testimoniato dal comunicato intitolato *Our preparations ahead of inauguration day*, con cui *Facebook-Meta* ha annunciato in modo generico le proprie intenzioni, l'11 gennaio 2020, reperibile al link: https://about.fb.com/news/2021/01/preparing-for-inauguration-day/ (14 maggio 2022). In quella sede, la piattaformadichiarava, ad esempio, «We are now removing content containing the phrase "stop the steal" under our Coordinating Harm policy from Facebook and Instagram» o «We are also restricting some features for people in the US based on signals such as repeat violations of our policies. These restrictions include blocking these accounts from creating live videos or creating an event, Group or Page». Cfr. anche il report R. AUGUSTINE, *A critique on content moderation on Facebook. A study based on "stop the steel" conspiracy campaign*, in *IJRCSS*, 21, 2021. Polemiche su un'accanimento da parte dei social media nei confronti della destra americana hanno, a partire da allora, assunto sempre più vigore, pur senza chiari elementi a supporto, v. A. GABBATT, *Claim of anti-conservative bias by social media firmsisbaseless, report finds*, The Guardian, 1 febbraio 2021.

cui sulle reti sociali si assiste al proliferare incontrollato di contenuti malevoli, un'attività che necessita sempre di "avvelenatori di pozzi", finanziatori interessati ad inquinare il dibattito pubblico[418]. Rimane più che dubbio che gli interventi censori che hanno colpito semplici persone fisiche fossero realmente necessari, o semplicemente che esistesse un chiaro nesso di causalità tra la condotta online di tali cittadini comuni e i problemi di ordine pubblico che si sono verificati. Né si può trascurare il valore di precedente di quanto avvenuto: account e pagine impegnati in un'attività massiva, professionale di contaminazione del discorso pubblico sono stati lasciati relativamente liberi di operare, fino al presentarsi di una situazione di estrema emergenza, cui si è risposto con un gran numero di provvedimenti *ex post* incidenti sulla libertà di espressione di persone comuni, spesso imprecisi e cui si è giunti, necessariamente, attribuendo all'automazione per mezzo dell'intelligenza artificiale un ruolo prevalente nella maggior parte di tali valutazioni[419].

Oltre a questi episodi - che è difficile non definire di censura di massa, almeno in alcune parti – ha suscitato particolari discussioni la decisione, da parte di diverse piattaforme, di limitare la possibilità di esprimersi attraverso di esse di soggetti normalmente aventi un ruolo privilegiato nel dibattito pubblico. La pratica, in realtà, è piuttosto risalente, e ha spesso suscitato polemiche e sospetti di parzialità nei confronti dei giganti del web, in primo luogo *Facebook*, da tempo accusato, ad esempio, di silenziare con una *content moderation* eccessiva attivisti curdi e palestinesi[420]. Ha cominciato, però, ad essere oggetto dell'attenzione di media e commentatori occidentali solo quando ha riguardato taluni partiti europeio personalità politiche di primo piano, tra cui lo stesso Donald Trump, al tempo capo dello stato in carica di una delle nazioni più influenti del pianeta[421]. Si tratta di valutazioni ponderate e discusse ai vertici delle compagnie coinvolte, che hanno segnato un nuovo corso nel ruolo dei giganti del web riguardo alla gestione del discorso pubblico. Complica ulteriormente il quadro il fatto che le piattaforme abbiano giustificato tali interventi sulla sola base delle proprie regole interne, qualificando come un comune rapporto contrattuale tra privati quello con gli utenti, anche quando essi siano soggetti del calibro di quelli coinvolti[422]. Un argomento portato avanti con pervicacia, come strategia difensiva, anche nei casi in cui l'esclusione dai servizi

[418]Cfr. N. GRINBERG ET AL., *Fake News on Twitter during the 2016 U.S. Presidential Election*, in *Science*, 363, 6425, 2019, p. 374–378; M. DEL VICARIO ET AL., *The Spreading of Misinformation Online*, in *Proceedings of the National Academy of Sciences*, 113, 3, p. 554–559; E. FERRARA ET AL., *The Rise of Social Bots*, in *Communications of the ACM*, 59, 7, p. 96–104; M. WOO, *How Online Misinformation Spreads*, in *Knowable Magazine-Annual Reviews*, 2021, https://bit.ly/3mOgZ9w (14 maggio 2022).

[419]Cfr. ancora F. C. MACKENZIE, *Fear the Reaper: how content moderation rules are enforced cit.*

[420] V. ad esempio Z. EL HAROUN, *Digital rights activists accuse Facebook of anti-palestinian bias*, Reuters, 3 novembre 2021; P. BOYLE, *Facebook censors support of Kurds*, in *Green Left Weekly*, 1 febbraio 2015.

[421]Il riferimento è, ovviamente, al rapporto complicato tra l'ex presidente USA e alcuni dei principali social media, culminato con la chiusura permanente dei suoi account, cfr. H. DENHAM, *These are the platformsthathave banned Trump and hisallies*, The Washington Post, 14 gennaio 2021.

[422] Si veda, ad esempio, il comunicato con cui Twitter, la prima piattaforma a oscurare l'account di Trump, ha motivato la scelta: *Permanent suspension of @realDonaldTrump*, 8 gennaio 2021, https://bit.ly/3zGr5kf (14 maggio 2022).

offerti dai giganti del web è stata portata di fronte a un giudice, come si dirà al paragrafo successivo.

4.4 La content moderation di fronte al giudice: i casi di deplatforming decisi dalle corti di Italia e Stati Uniti come dimostrazione dell'incoerenza dell'attuale statuto giuridico dei social media

Due dei più significativi casi di c.d. *de platforming* (così è ormai comune riferirsi all'esclusione da una determinata piattaforma per la violazione delle condizioni d'utilizzo)[423] su cui è stata chiamata ad esprimersi un'autorità giudiziaria sono stati discussi in Italia, e in particolare di fronte al Tribunale di Roma nei primi mesi del 2021, in seguito alla decisione del *social network* Facebook di oscurare gli account ufficiali di due partiti politici di destra radicale, CasaPound e Forza Nuova, e dei loro principali leader[424]. Significativo è che i ricorsi presentati in via cautelare ex art. 700 c.p.c. da entrambi i partiti abbiano avuto esiti antitetici, portando, nel primo caso, alla riattivazione delle pagine eliminate dalla piattaforma, poi confermata in sede di reclamo, e alla conferma del *ban*nel secondo[425]. Da quanto si evince dai provvedimenti conclusivi, i due movimenti politici avrebbero affrontato il giudizio con prospettazioni molto simili, argomentando che la chiusura degli account ledeva diritti di rango costituzionale come la libertà d'azione o le prerogative riconosciute ai partiti, e giustificando la necessità di un decreto d'urgenza con la posizione di preminenza assunta nel dibattito pubblico dai social network, che portava a considerare un grave *vulnus* l'esclusione da essi,

[423] Il termine è approdato addirittura su Wikipedia, cfr. https://en.wikipedia.org/wiki/Deplatforming (20 maggio 2022).
[424] M. PENNISI, *CasaPound e Forza Nuova rimossi definitivamente da Facebook e Instagram: «Diffondono odio»*, in *Corriere della Sera (online)*, 9 settembre 2019; S. COSTANTINI, *I social sgomberano Casa Pound*, in *La Repubblica*, 10 settembre 2019; G. LONGO, *Facebook chiude i gruppi CasaPound e Forza Nuova*, in *La Stampa*, 10 settembre 2019.
[425] A decidere i due procedimenti sono stati i citati provvedimenti giudiziari Tribunale di Roma – sez. imprese, ord. 12 dicembre 2019, confermato in sede di reclamo da Tribunale di Roma – XVII sez. civile, 29 aprile 2020, per quanto riguadaCasaPound, e Tribunale di Roma – sez. diritti della persona e immigrazione, ord. 23 febbraio 2020, per quanto riguarda Forza Nuova. Per alcuni commenti in letteratura si rimanda a C. MELZI D'ERIL, G. E. VIGEVANI, *Odio in rete e rimozione delle pagine Facebook: giudice che vai, soluzione che trovi*, in *Il Sole 24 Ore*, 27 febbraio 2020; L. RINALDI, *Le piattaforme tra diritto pubblico e diritto privato cit.*; A. QUARTA, *Disattivazione della pagina Facebook. Il caso CasaPound tra diritto dei contratti e bilanciamento dei diritti*, in *Danno e responsabilità*, 4/2020, 489 ss.; A. VIGORITO, *Piattaforme digitali e "politicalspeech": dal caso Facebook-CasaPound alla vicenda Trump-Twitter*, in *giustiziacivile.com*, 11/2020,16 ss.; P. VILLASCHI, *Facebook come la RAI? Note a margine dell'ordinanza del Tribunale di Roma del 12.12.2019 sul caso CasaPound c. Facebook*, in *Osservatorio AIC*, 2/2020, 430 ss.; P. FALLETTA, *Controlli e responsabilità dei "social network" sui discorsi d'odio "online"*, in *MediaLaws*, 1/2020,146 ss.; A. GOLIA JR., *L'antifascismo della Costituzione italiana alla prova degli spazi giuridici digitali. Considerazioni su partecipazione politica, libertà di espressione "online" e democrazia (non) protetta in "CasaPound c. Facebook" e "Forza Nuova c. Facebook"*, in *Federalismi.it*, 18/2020,134 ss.; S. PIVA, *Facebook è un servizio pubblico? La controversia su CasaPound risolleva la "quaestio" dell'inquadramento giuridico dei "social network"*, in *diritti fondamentali.it*, 2/2020,1192 ss. O. GRANDINETTI, *Facebook vs. Casa Pound e Forza Nuova, ovvero la disattivazione di pagine social e le insidie della disciplina multilivello dei diritti fondamentali*, in *Media Laws*, 1/2021, 173 ss; P. ZICCHITTU, *I movimenti "antisistema" nell'agorà digitale: alcune tendenze recenti*, in *giurcost.org*, 5 marzo 2020 e *La libertà di espressione dei partiti politici nello spazio pubblico digitale: alcuni spunti di attualità*, in *MediaLaws*, 2/2021, p. 2 ss.

anche solo per pochi giorni[426]. La piattaforma, dal canto suo, ha impostato la difesa in ambo i casi sul menzionato argomento della semplice messa in atto di condizioni contrattuali accettate dagli utenti con l'iscrizione, giungendo ad affermare in atti, non senza una certa *hybris*: «La circostanza che si tratti di un'organizzazione proibita o meno secondo la legge italiana non assume alcuna rilevanza»[427]. La contradditorietà delle due soluzioni, peraltro elaborate dallo stesso Tribunale (cambiava, ovviamente, la persona fisica del giudice) è sintomatica della complessità di questo genere di questioni e del delicato bilanciamento che sottendono, cui l'attuale quadro giuridico risponde, probabilmente, in modo troppo incerto, generando palesi incoerenze come quella in esame. Ciò che preme sottolineare, però, è che, nonostante la radicale diversità, i provvedimenti dei giudici di entrambi i procedimenti concordano su un punto centrale: le piattaforme hanno assunto un ruolo peculiare nel mercato dell'informazione e nel dibattito pubblico, da cui non possono non derivare conseguenze. Da questa constatazione discendono, a seconda della pronuncia, conseguenze opposte. Nel caso di CasaPound, si sancisce che il valore che i *social network* oggi rappresentano per la libertà d'espressione e il pluralismo informativo funge da limite di diritto pubblico all'autonomia privata, e, dunque, l'esclusione dalla piattaforma non può basarsi solo su poche regole interne, vagamente formulate[428]. Nel caso di Forza Nuova, invece, la posizione di preminenza assunta nel dibattito collettivo è assunta dal giudice come argomento per giustificare la condotta di Facebook, sostenendo che la permanenza sulla piattaforma del partito avrebbe rappresentato una lesione di diverse Convenzioni internazionali e della legislazione nazionale contro la discriminazione razziale, della normativa sul divieto di ricostituzione del partito fascista e di diverse fonti di *soft-law*[429]. Facebook, dunque, pare onerato del compito di interpretare ed applicare, in prima battuta, un quadro normativo variegato e complesso, non concepito per lo scenario digitale e che non prevede, in ogni caso, l'obbligo esplicito di oscurare account o censurare

[426] Si vedano le line difensive dei due partiti politici ricostruite in Tribunale di Roma – sez. imprese, ord. 12 dicembre 2019 cit., 4-5 e Tribunale di Roma – sez. diritti della persona e immigrazione, ord. 23 febbraio 2020 cit., 15-17.

[427] Riporta queste parole, estratte dalla memoria di costituzione della società resistente nel giudizio di primo grado, l'ordinanza del Tribunale di Roma – XVII sez. civile, 29 aprile 2020, che ha chiuso il giudizio di reclamo ex art. 669-terdecies c.p.c. incardinato da Facebook contro l'ordinanza di accoglimento del ricorso di CasaPound.

[428] Nell'ordinanza del Tribunale di Roma – sez. imprese,12 dicembre 2019 cit., infatti, il Giudice cautelare afferma: «È infatti evidente il rilievo preminente assunto dal servizio di Facebook (o di altri social network ad esso collegati) con riferimento all'attuazione di principi cardine essenziali dell'ordinamento come quello del pluralismo dei partiti politici (art. 49 Cost.), al punto che il soggetto che non è presente su Facebook è di fatto escluso (o fortemente limitato) dal dibattito politico italiano […]. Ne deriva che il rapporto tra Facebook e l'utente che intenda registrarsi al servizio (o con l'utente già abilitato al servizio come nel caso in esame) non è assimilabile al rapporto tra due soggetti privati qualsiasi in quanto una delle parti, appunto Facebook, ricopre una speciale posizione: tale speciale posizione comporta che Facebook, nella contrattazione con gli utenti, debba strettamente attenersi al rispetto dei principi costituzionali e ordinamentali».

[429] Il provvedimento prende in consdierazione, tra le atre cose, la Dichiarazione universale dei diritti dell'uomo del 1948, il Patto per i diritti civili e politici e la Convenzione di New York sull'eliminazione di tutte le forme di discriminazione razziale del 1966, la Convenzione EDU, la Carta dei diritti fondamentali dell'Unione Europea, le c.d. leggi Mancino (Legge n. 205 del 25 giugno 1993) e Scelba (Legga n. 645 del 20 giugno 1952), cfr. Tribunale di Roma – sez. diritti della persona e imm., ord. 23 febbraio 2020 cit., p. 1-14.

contenuti[430]. Un compitoche va ben al di là dell'applicazione di determinate condizioni contrattuali nell'ambito di un negozio di diritto privato.

Anche l'esclusione di Trump dalle piattaforme, inaugurata da Twitter tra il 6 e il 7 gennaio 2021, in conseguenza dei fatti di Capitol Hill, con una decisione poi replicata, in pochi giorni, da tutti gli altri principali operatori,è stata portata di fronte a un giudice[431]. Nel 2021, infatti, l'ormai ex presidente ha avviato un'azione civile nei confronti di Twitter, per ottenere la condanna della piattaforma alla riapertura del suo account[432]. La vicenda giudiziaria più significativa che ha coinvolto Trump e le piattaforme, eTwitter in particolare, risale, però, ad alcuni anni prima. A partire dal 2017, infatti, l'allora presidente degli Stati Uniti è stato coinvolto in un giudizio, incardinato di fronte alle corti dello stato di New York da un gruppo di sette cittadini americani e dal *Knight First Amendment Institute*, un think-tank in materia di libertà d'espressione afferente alla

[430]Infatti, l'ordinanza Tribunale di Roma – sez. diritti della persona e imm., 23 febbraio 2020 cit., svolti i riferimenti al diritto nazionale e internazionale appena visti (cfr. nota precedente) prosegue analizzando le clausole contrattuali invocate da Facebook a sostegno della propria decisione. Vengono in gioco, in particolare, i punti 1 (*Servizi offerti da Facebook*), 3.2 (*Elementi condivisibili e condotte autorizzate su Facebook*) e 4.2 (*Sospensione o chiusura dell'account*) delle Condizioni d'Uso, https://www.facebook.com/terms/ (14 maggio 2022) e il titolo *Persone e organizzazioni pericolose* degli Standard della Comunità, https://transparency.fb.com/policies/community-standards/ (14 maggio 2022). Il provvedimento, poi, esamina una lunga serie di contenuti diffusi, negli anni precedenti, dagli account del movimento politico e dei suoi membri, argomentandone l'illiceità ai sensi delle regole interne di Facebook e della normativa citata in precedenza. La piattaforma, dunque, pare onerata dell'*enforcement*in prima battuta di tali norme, internazionali e nazionali, di *hard law* e *soft law*, e delle delicate valutazioni che ne conseguono, in virtù della posizione di preminenza assunta nel mercato dell'informazione. Ciò nonostante non si tratti, in larga maggioranza, di regole relative alla circolazione di contenuti online, e risultino, dunque, applicabili al contesto in esame solamente con sforzi interpretativi talora notevoli. In particolare, nel provvedimento, a pagina 43, il Giudice cautelare afferma: «I contenuti, che inizialmente erano stati rimossi e poi a fronte della reiterata violazione hanno comportato la disattivazione degli account dei singoli ricorrenti e delle pagine da loro amministrate tutte ricollegabili a Forza Nuova, sono illeciti da numerosi punti di vista. Non solo violano le condizioni contrattuali, ma sono illeciti in base a tutto il complesso sistema normativo di cui si è detto all'inizio, con la vasta giurisprudenza nazionale e sovranazionale citata. Facebook non solo poteva risolvere il contratto grazie alle clausole contrattuali accettate al momento della sua conclusione, ma aveva il dovere legale di rimuovere i contenuti, una volta venutone a conoscenza, rischiando altrimenti di incorrere in responsabilità (si veda la sentenza della CGUE sopra citata e la direttiva CE in materia), dovere imposto anche dal codice di condotta sottoscritto con la Commissione Europea». La sentenza della CGUE cui si fa riferimento è C-18/18 Glawischnig-Piesczek, che ha stabilito che il giudice di uno Stato membro può ordinare a Facebook la rimozione di un contenuto illecito e di altri che sembrino equivalenti, ribadendo al contempo che sui fornitori di servizi di hosting non gravano obblighi generali di sorveglianza. Preme evidenziare, però, che nel caso giunto di fronta alla Corte di Giustizia dell'Unione Europea l'originaria valutazione di illiceità del contenuto proveniva dall'autorità giudiziaria e non era delegata alla piattaforma, onerata unicamente del reperimento di altri materiali ad esso accostabili senza sforzi interpretativi significativi. Per un commento cfr. M. MONTI, *La Corte di giustizia, la direttiva e-commerce e il controllo contenutistico online: le implicazioni della decisione C 18-18 sul discorso pubblico online e sul ruolo di Facebook*, in *MediaLaws*, 3, 2019, 1 ss.

[431]Cfr. ad es. H. DENHAM, *These are the platforms that have banned Trump and his allies*, in *The Washington Post*, 14 gennaio 2021. Le piattaforme hanno giustificato la decisione facendo riferimento pressochè esclusivo alle proprie regole interne: cfr. ad es. *Permanent suspension of @realDonaldTrump*, 8 gennaio 2021, https://blog.twitter.com/en_us/topics/company/2020/suspension (20 maggio 2022). Per quanto riguarda *Facebook*, questa impostazione è stata poi adottata anche dall'Oversight Board della piattaforma, che ha confermato il *ban* dell'ex presidente, cfr. Facebook Oversight Board, Case decision 2021-001-FB-FBR, 5 maggio 2021, https://www.oversightboard.com/decision/FB-691QAMHJ (20 maggio 2022).

[432]Cfr. ad es. K. LYONS, *Trump suesto reinstate his Twitter account*, The Verge, 2 ottobre 2021. Per il ricorso presentato da Trump v. District Court N.D., California, *Trump v. Twitter, Inc., No. 3:21-cv-08378 – Compliant*, 7 luglio 2022, in *CourtListener.com*, https://bit.ly/3tSFjel (20 maggio 2022).

Columbia University[433]. I sette privati lamentavano di essere stati tutti "bloccati" dall'account ufficiale della presidenza Trump su Twitter dopo aver espresso dissenso riguardo alle azioni del capo dello stato nella sezione commenti (la funzione "blocco" di Twitter permette di impedire a un utente di visualizzare, commentare o condividere i contenuti diffusi col proprio account). Essi, in buona sostanza, sostenevano che il comportamento del presidente violasse la prerogativa del *free speech*ad essi garantita dal Primo Emendamento, impedendo loro di commentare e criticare i post dell'account presidenziale. Il *Knight First AmendmentInstitute*, dal canto suo, argomentava la lesione della propria libertà negativa d'informazione, poiché il *ban* dei sette utenti impediva di leggere e conoscere quanto, in caso contrario, essi avrebbero scritto in reazione ai contenuti diffusi da Trump[434]. La prospettazione dei ricorrenti è stata accolta sia in primo grado che in appello, con l'argomento che l'utilizzo dell'account Twitter come un mezzo ufficiale di comunicazione da parte di Trump ne aveva fatto uno spazio pubblico digitale soggetto alla *public forum doctrine* e in cui i poteri pubblici, *in primis* il presidente, non potevano discriminare in base al contenuto le opinioni che vi venivano espresse dai cittadini[435].

La causa è stata poi dichiarata *moot*, e dunque cancellata dal ruolo perché non più attuale, dalla Corte Suprema, che ha discusso il caso solo quando Trump, ormai, non era più il presidente in carica[436]. Sono di particolare interesse, però, le considerazioni espresse in tale sede, e in particolare in un'estesa *concurring opinion* di Justice Clarence Thomas, il giudice con la più alta anzianità di servizio, riguardo allo statuto giuridico delle piattaforme[437]. Justice Thomas, infatti, evidenzia la necessità di elaborare una concezione degli operatori di internet che, pur rispettandone la natura di enti di diritto privato, sia ingrado di rispondere in maniera efficente alle problematiche sempre più urgenti in materia dilibertà d'espressione e controllo del discorso pubblico che si presentano al loro

[433] Cfr. https://knightcolumbia.org/page/about-the-knight-institute (20 maggio 2022).

[434] Per l'atto introduttivo del giudizio cfr. District Court S.D., New York, *Knight First Amendment Institute v. Trump, No. 1:17-cv-05205 – Compliant*, 11 luglio 2017, in *CourtListener.com*, https://bit.ly/3xnV6AM(20 maggio 2022).

[435]Cfr., per la pronuncia di primo grado, District Court S.D., New York, *Knight First Amendment Institute v. Trump, No. 1:17-cv-05205 – Order on motion for Summary Judgment*, 23 maggio 2018, in *CourtListener.com*, https://bit.ly/3fmoalI (20 luglio 2021) e, per la sentenzad'appello, U.S. Court of Appeals 2nd Circuit, *Knight First Amendment Institute v. Trump, No. 18-1691-cv*, 9 luglio 2019, in *Justia US Law*, https://law.justia.com/cases/federal/appellate-courts/ca2/18-1691/18-1691-2019-07-09.html(20 maggio 2022). Per deicommentisivedano*Recent case: Knight First Amendment Institute at Columbia University v. Trump*, in *Harvard Law Review Blog,* 3 giugno 2019, https://bit.ly/3rGC1bA (20 maggio 2022); J. ROBERTS, *Trump, Twitter, and the First Amendment*, in *Alternative Law Journal*, 44, 3/2019, 207 ss.; con toniaspramentecriticiL. BEAUSOLEIL, *Is trolling Trump a right or a privilege? The erroneous finding in Knight First Amendment Institute at Columbia University v. Trump*, in *Boston College Law Review*, 60, 9/2019, 31 ss. Riguardo alla*public forum doctrine*, la sua prima enunciazione risale al 1939, nell'opinione di Justice O.J. Roberts in *Hague v. Committee for Industrial Organization*; In letteratura possono indicarsi *ex multis* D.L. HUDSON JR., *Public forum doctrine*, in *The First Amendment Encyclopedia*, 2020, https://mtsu.edu/first-amendment/article/824/public-forum-doctrine (21 luglio 2021);R.A. HORNING, *The first amendment right to a public forum*, in *Duke Law Journal*, 1969, 931 ss.; R.C. POST, *Between governance and management: the history and theory of the public forum*, in *UCLA Law Review*, 34/1987, 1713 ss.

[436]U.S. Supreme Court, *Biden v. Knight First Amendment Institute at Columbia University*, 5 aprile 2021.

[437] Per un commento in letteratura v. M. MONTI, *La Corte Suprema statunitense e il potere delle piattaforme digitali: considerazioni sulla privatizzazione della censura a partire da una concurring opinion*, in *DPCE online*, 1/2021.

interno, di cui il *ban* del presidente Trump è un paradigmatico esempio[438]. Dunque, anche i giudici americani, al pari di quelli italiani, ritengono urgente, a tutela dei diritti fondamentali coinvolti nel loro funzionamento, ripensare l'inquadramento giuridico delle piattaforme. Ciò è ancor più significativo in un ordinamento come quello statunitense, in cui l'efficacia unicamente verticale del *Bill of Rights*(e dunque anche della protezione del *free speech*) non è pressochè mai stata messa in discussione e appare molto più problematico che negli ordinamenti europei affermare che la tutela della libertà di manifestazione del pensiero possa rappresentare un limite all'autonomia negoziale di società di diritto privato come le piattaforme[439].

L'indagine comparata, oltre a rivelare queste analogie, mette in luce come l'attuale strumentario giuridico, da ambo i lati dell'Atlantico, dia luogo a palesi incongruenze, dimostrandosi deficitario anche da tale punto di vista. Se appare quasi paradossale, infatti, che il Tribunale di Roma sia giunto a soluzioni antitetiche in due casi pressoché identici, in materia di censura di partiti estremisti sui social network, non può non suscitare una riflessione che Twitter abbia disinvoltamente chiuso come un qualunque spazio fisico privato, dichiarando di agire sulla sola base delle proprie regole interne, l'account di Donald Trump, la cui sezione commenti è stata, però, allo stesso tempo considerata un *public forum*, in cui il presidente non può interferire col libero scambio di idee e opinioni, dai giudici americani[440]. L'incoerenza non è che acuita dalla circostanza che la mezionata causa contro Twitter intentata da Trump nel 2021 per la riapertura dell'account sia stata dichiarata *dismessed* nel maggio 2022 dal Giudice Federale della California, in seguito a una valutazione sommaria e senza la celebrazione di udienze, con uno stringato *order* in cui l'argomento principale è rappresentato dalla natura totalmente privata di Twitter e delle sue decisioni, che non possono considerarsi la manifestazione, nemmeno indiretta, di un potere dello stato. Il privato che ne venisse pregiudicato non può, allora, appellarsi alle garanzie del Primo Emendamento, applicabile solo nei confronti dei poteri pubblici o degli operatori privati che, nel concreto, si riducano a semplice strumento d'azione di questi, in base alla c.d. *state actiondoctrine*[441].

Prima del *ban*, da parte di *Facebook*, dei partiti italiani Forza Nuova e CasaPound, l'unico precedente significativo era rappresentato dall'esclusione dalle piattaforme dei leader e della pagina

[438] Cfr. ancora U.S. Supreme Court, *Biden v. Knight First Amendment Institute at Columbia University*, 5 aprile 2021.

[439]Sulla concezione del *free speech* negli Stati Uniti si rimanda ai già citati G. R. STONE, L. C. BOLLINGER, *The free speech century cit.*; F. ABRAMS, *The soul of the first amendment cit.*; Z. CHAFEE, *Free speech in the United States cit.*; M.ROSENFELD, A. SAJO, *Spreading liberal constitutionalism: an inquiry into the fate of free speech rights in new democracies cit.*; O. POLLICINO, *La prospettiva costituzionale sulla libertà d'espressione nell'era di internet cit.*

[440] Lo rileva anche lo stesso Justice Thomas, nella menzionata *concurring opinion*, par. 1: «it seems rather odd to say that something is a government forum when a private company has unrestricted authority to do away with it».

[441]Cfr. District Court N.D., California, *Trump v. Twitter, Inc., No. 3:21-cv-08378 – Order remotion to dismiss cit.*, p. 3-13. Sulla *state action doctrine* cfr., *ex multis,* S. JAGGI, *State action doctrine*, in *Max Planck Encyclopedia of Comparative Constitutional Law*, 2017, https://oxcon.ouplaw.com/view/10.1093/law-mpeccol/law-mpeccol-e473 (20 luglio 2021); T. PERETTI, *Constructing the State Action Doctrine, 1940–1990*, in *Law & Social Inquiry*, 35, 2/2010, 273 ss.

ufficiale del movimento politico estremista britannico *British First*, in tal caso, però, preceduta dalla messa fuori legge del movimento e dalla condanna dei politici interessati per crimini d'odio[442]. Successivamente a quanto avvenuto con Donald Trump, in ogni caso, la pratica sembra essere divenuta più comune, e *Facebook*, nel corso del 2020 e del 2021, ha messo al bando due politici israeliani di estrema destra e il partito polacco *Konfederacja*[443]. È lecito aspettarsi, dunque, che vicende di *deplatforming* tornino in tempi brevi di fronte ai giudici, e ciò rende particolarmente urgente l'elaborazione di uno statuto giuridico delle piattaforme solido e coerente: un'operazione che le corti di Italia e Stati Uniti hanno mostrato essere particolarmente complessa e foriera di interrogativi da risolvere.

5. Adeguare un quadro giuridico non più attuale: le ipotesi di regolazione attualmente in discussione e alcuni spunti *de iure condendo* per una protezione effettiva della libera manifestazione del pensiero sui *social media*

Appurato che l'attuale quadro giuridico appare, da molti punti di vista, insufficiente per una regolazione soddisfacente dell'attività di *contentmoderation*, pare opportuno, a conclusione della disamina dell'argomento, indagare alcune possibili prospettive *de iure condendo*. Come già visto, l'intero sistema tutt'oggi si regge sul principio, di creazione non più recentissima, dell'esenzione degli intermediari da una responsabilità generalizzata per i contenuti diffusi dagli utenti, a patto che non esercitino su di essi una condotta che possa considerarsi "attiva" (dal punto di vista della creazione o della manipolazione). L'obbligo di rimuovere un contenuto illecito, e la conseguente responsabilità azionabile in giudizio, sorgono solamente qualora la piattaforma venga in qualunque modo a conoscenza di tale materiale. Il principio è stato di recente messo in forte discussione, sia in Europa che negli Stati Uniti: tuttavia, non pare agevole identificare i vantaggi cui il suo abbandono potrebbe condurre[444]. Infatti, pare difficile imputare alla *liability exemption* gran parte dei problemi connessi alla moderazione dei contenuti affrontati fin qui. In primo luogo, perché molti di essi derivano dalla scarsa trasparenza, dall'incoerenza e dai limiti tecnici dell'attività di moderazione messa in atto dalle piattaforme al di fuori di ogni obbligo di legge[445]. In secondo luogo, perché il passaggio a un regime di responsabilità generalizzata non risolverebbe, probabilmente, il problema

[442]HERN, K. RAWLINSON, *Facebook bans Britain First and its leaders*, in *The Guardian (online)*, 14 marzo 2018; D. BROWN, *Britain First defies ban on Facebook*, in *The Times*, 15 marzo 2018.
[443]T. STAFF, *Facebook bans far-right political figures Marzel, Ben Ari from its platforms*, The Times of Israel, 11 agosto 2021; *Polish far-right party banned from Facebook over alledged COVID disinformation*, Euronews, 6 gennaio 2022.
[444] Per alcune critiche al principio, cfr. ancora M. D. SMITH, M. VAN ALSTYNE, *It's time to update Section 230 cit.*; J. ADETUNJI, *Tech giants need to take more responsibility cit.*; M. REARDON, *Section 230: how it shields Facebook and why Congress wants changes cit*.
[445]J. C. YORK, C. MCSHERRY, *Content Moderation Is Broken. Let Us Count the Ways cit.*

più delicato: la delicatezza e l'intrinseca politicità di molte valutazioni relative alla *contentmoderation*, e in particolare di quelle che interferiscono con le attività di soggetti che rivestono un ruolo di particolare rilievo nel dibattito pubblico, come partiti o titolari di cariche elettive.Inoltre, gli svantaggi della costruzione di un obbligo legale di monitoraggio in capo alle piattaforme, in termini di eccessiva *compliance*censoria delle piattaforme e conseguente*chillingeffect*degli utenti,appaiono, al contrario dei vantaggi, molto chiari [446]. In particolare se si tiene conto che il semplice timore di perdere il privilegio della *liabilityexemption*, ed essere quindi gravate di responsabilità, è stato tra le spinte principali per lo sviluppo di sistemi di moderazione dei contenuti sempre più stringenti, che negli ultimi anni si sono spinti, come abbiamo visto, ben oltre il semplice filtro dei materiali palesemente illegali o in grado di turbare fortemente il pubblico. È ragionevole presumere, dunque, che il passaggio a un regime che non prevedesse l'esenzione porterebbe allo sviluppo, da parte dei *social media*, di un regime censorio di inedita intensità.

Allo stesso tempo, non si può non sottolineare comel'attuale quadro giuridico appaia ormai, almeno in parte, inadeguato all'evoluzione del mercato dei servizi internet. L'idea di una passività delle piattaforme rispetto ai materiali che gli utenti diffondono attraverso di esse, alla base dell'esenzione da responsabilità, pare sempre meno credibile: le attività di indicizzazione e profilazione, la vendita di spazi pubblicitari personalizzati e il resto delle operazioni svolte dagli intermediari sui contenuti – non ultima la stessa moderazione – somigliano sempre più all'esercizio di un controllo effettivo[447]. In più, la *liability exemption* gioca senza dubbio un ruolo centrale in una delle questioni più ampie e urgenti legate all'attuale struttura della *content moderation*: il già analizzato protagonismo, pressochè incontrastato, delle piattaforme, che dichiarano esplicitamente di agire solo sulla base delle proprie regole interne, spesso al di fuori di un preciso *framework* legale o nel quadro di generici strumenti di auto o co-regolazione che rappresentano, nella pratica, una delega in bianco[448]. L'attuale quadro normativo dovrebbe, dunque, essere migliorato, ma non stravolto nei principi di fondo. Cercando di abbozzare una soluzione, la *liability exemption* dovrebbe essere accompagnata da stringenti obblighi procedurali in capo alle piattaforme, imponendo loro di

[446]Cfr. ad es. G. F. FROSIO, *Reforming intermediary liability in the platform economy cit.* e *Why keep a dog and bark yourself? cit.*; S. STALLA-BOURDILLON, *Liability exemptions wanted! cit.*; R. DARA, *Intermediary Liability in India: Chilling Effects on Free Expression on the* Internet, 2011, http://dx.doi.org/10.2139/ssrn.2038214.

[447]T. GILLESPIE, *Platforms are not intermediaries*, in *Georgetown Law Technology Review*, 2, 2, p. 198-216; L. DE NARDIS, A. M. HACKL, *Internet governance by social media platforms*, in *Telecommunications Policy*, 39, 9, 2015, p. 761-770.

[448] È solo l'assenza di obblighi legali vincolanti a permettere alle piattaforme di regolare l'attività di moderazione con *corpus* normativi sempre più ampi di diritto privato, di fatto autoprodotti. Cfr. E. CREMONA, *Le nuove tecnologie oltre la "grande dicotomia" tra pubblico e privato*, in *Gruppo di Pisa. Dibattito aperto sul Diritto e la Giustizia costituzionale*, Quad. Monografico n. 3 - fascicolo 2, 2021, p. 681 ss., e ancora D. KELLER, *Making Google the censor cit.*; K. KLONICK, *The new governors cit.*; M. MONTI, *Privatizzazione della censura e internet platforms cit.*

mettere in atto il massimo standard di qualità esigibile nell'attività di moderazione al fine di perseguire due obiettivi: la più alta precisione ottenibile nel filtro di contenuti illeciti, con il minor impatto possibile sulla libertà di espressione degli utenti, oscurando materiali soltanto ove necessario, concentrando gli sforzi sui superdiffusori di contenuti malevoli e garantendo il diritto di intervento dei soggetti coinvolti. Si tratta, peraltro, di una soluzione che porta con sé la necessità di ripensare le attuale modalità con cui la tecnologia, e in particolare l'intelligenza artificiale, è utilizzata nell'attività di *content moderation*, come visto un altro dei problemi più urgenti sollevati dall'attuale situazione.

A tal proposito, deve evidenziarsi, ancora una volta, che le dimensioni che i servizi dell'internet 2.0 hanno assunto rendono inevitabile il ricorso all'automazione, posta la quantità di contenuti diffusi ogni giorno sulle piattaforme. Inoltre, al netto dei limiti menzionati in precedenza, l'efficacia su larga scala della moderazione svolta da algoritmi, spesso basati sull'intelligenza artificiale, è indubbia, e la tecnologia ha un ruolo insostituibile nell'impedire che contenuti estremamente disturbanti, come immagini oscene o ritraenti delitti efferati, giungano agli occhi di milioni di persone[449]. Eventuali modifiche dell'assetto esistente dovrebbero puntare a cambiare le attuali modalità di interazione tra elemento artificiale e umano nelle decisioni attinenti alla moderazione dei contenuti, e non a limitare il ruolo della tecnologia. In particolare, un eventuale intervento normativo dovrebbe avere, come già detto, l'obiettivo di spingere le piattaforme a permetterel'intervento del soggetto che subisce il provvedimento censorio, nelle sue varie forme[450].Le attuali pratiche delle piattaforme sono, infatti, estremamente variegate e vanno dalla rimozione di un determinato contenuto, al *flagging* con determinati avvisi (come quello della possibile presenza di *fake news*) o alla radicale espulsione dai servizi[451]. In tale ottica, il responsabile del contenuto moderato dovrebbe essere in grado di sottoporre la decisione a un operatore umano per una revisione in tempi ragionevoli, avendo la possibilità di difendere succintamente le proprie ragioni e di ricevere una motivazione in caso di conferma del provvedimento censorio. Il problema della mole potenzialmente molto elevata di istanze dovrebbe essere risolto, in primo luogo, con l'impiego di un volume di risorse proporzionato da parte delle

[449] Il rinvio è ancora a R. GORWA, R. BINNS, C. KATZENBACH, *Algorithmic content moderation cit.*; E. LLANSÒ, J. VON HOBOKEN, P. LEERSSEN, J. HARAMBAM, *Artificial intelligence, content moderation and freedom of expression cit.*; F. C. MACKENZIE, *Fear the Reaper: how content moderation rules are enforced on social media cit*.

[450]Fanno riferimento, da prospettive diverse, al diritto a un appello – rivolto a uno o più esseri umani - contro la moderazione di un determinato contenuto S. QUINTARELLI, *Content moderation: i rimedi tecnici cit.*, 147-148; J. C. YORK, C. MCSHERRY, *Content ModerationIsBroken cit.*; N. DUARTE, E. LLANSO, A. LOUP, *Mixed messages? The limits of automated social media content analysis cit.*; P. BARRETT, *Who moderates the social media giants?*, Report - NYU Stern, 2020, https://bit.ly/3txnHEj (15 maggio 2022).

[451]Per una panoramica, vedi in particolare E. LLANSÒ, J. VON HOBOKEN, P. LEERSSEN, J. HARAMBAM, *Artificial intelligence, content moderation and freedom of expression cit.*, p. 14-18.

piattaforme (non pare irragionevole, posti gli attuali astronomici margini di guadagno)[452]. Non è difficile, inoltre, immaginare strategie per razionalizzare il lavoro di questi revisori umani: dovrebbero avere la priorità, in primo luogo, i contenuti moderati solo in seguito alla segnalazione da parte degli altri utenti della piattaforma. Essi, infatti, hanno superato il filtro preliminare messo in atto da quest'ultima con modalità automatizzate, che impedisce sul nascere la condivisione di un gran numero di contenuti illeciti. La loro revisione, dunque, appare urgente per due motivi: esiste un chiaro rischio di segnalazioni interessate o comunque non fondate da parte degli altri utenti, e dunque di un possibile *vulnus* alle prerogative del soggetto coinvolto; potrebbe trattarsi di contenuti malevoli in grado di superare un primo filtro automatizzato, o che magari ha già coinvolto l'intervento di esseri umani, e che dunque appaiono, da un lato, particolarmente preziosi per la piattaforma stessa, pochè in grado di evidenziare le criticità del sistema di moderazione, e, dall'altro, caratterizzati, con maggior probabilità, da un elevato tasso di ambiguità e di opinabilità della decisione, che rende un riesame particolarmente opportuno. In secondo luogo, dovrebbe darsi preferenza alle richieste di revisione provenienti da soggetti mai incorsi in sanzioni connesse alla *content moderation* della piattaforma, posta l'evidente maggior probabilità che tali istanze non si rivelino pretestuose, contribuendo indirettamente ad identificare disfunzioni del sistema di moderazione. Parallelamente, soggetti "recidivi" e superdiffusori di contenuti malevoli dovrebbero ricevere un trattamento sfavorevole dal punto di vista sanzionatorio, fino all'esclusione generalizzata dalla piattaforma. Molti degli strumenti elencati, in ogni caso, sono già implementati da varie piattaforme, compresa la possibilità di revisione delle decisioni di moderazione, pur con modalità che appaiono molto meno trasparenti e garantiste di quanto appena sinteticamente proposto. Questo, peraltro, pare un argomento ulteriore a favore della fattibilità tecnica di questi miglioramenti[453].

Si tratta di soluzioni che sembra condividere, nelle linee generali, anche una recente proposta legislativa della Commissione europea, presentata nel dicembre 2020 e volta a introdurre, con Regolamento, un *Digital service act* destinato a superare l'attuale Direttiva *e-commerce*[454]. Il testo,

[452] Aumentare radicalmente il numero dei moderatori umani (e migliorare le loro condizioni di lavoro) d'altronde, è tra le raccomandazioni più comuni che gli studi sul tema rivolgono alle piattaforme: cfr. ad es. P. BARRETT, *Who moderates the social media giants? cit.*, p. 24 ss.

[453] S. SINGH, *Everything in Moderation. An Analysis of How Internet Platforms Are Using Artificial Intelligence to Moderate User-Generated Content*, New America – Open Technology Institute, 2019, https://bit.ly/2xZhPqu (14 maggio 2022); P. BARRETT, *Who moderates the social media giants? cit.*, p. 7-19 ss.; F. C. MACKENZIE, *Fear the Reaper: how content moderation rules are enforced cit*.

[454] COMMISSIONE EUROPEA, *Proposta di Regolamento al Parlamento Europeo e al Consiglio relativo a un mercato unico dei servizi digitali (legge sui servizi digitali) e che modifica la direttiva 2000/31/CE*, 15 dicembre 2020, COM(2020) 825 final. In letteratura, per alcuni primi commenti, cfr. M. D. COLE, C. ETTELDORF, U. CARSTEN, *Updating the Rules for Online Content Dissemination: Legislative Options of the European Union and the Digital Services Act Proposal*, 2022, https://www.nomos-elibrary.de/index.php?doi=10.5771/9783748925934 (14 maggio 2022); G. FINOCCHIARO, *Digital Services Act: la ridefinizione della limitata responsabilità del provider e il ruolo*

opportunamente, distingue gli intermediari di internet in varie categorie, in base a funzioni e dimensioni, differenziando le condotte che dovranno mettere in atto se la proposta sarà approvata. Rimane ferma l'esclusione di una responsabilità generalizzata delle piattaforme per i contenuti diffusi dagli utenti e il sistema dovrebbe continuare a reggersi, quindi, sul principio della *liabilityexemption*. Nello specifico, il testo distingue tra "intermediari puri", che forniscono servizi di c.d. *mere conduit*o *chachin*g e dunque non immagazzinano a lungo termine i contenuti degli utenti[455], *hosting services*, che li immagazzino senza diffonderli (come i servizi *cloud*)[456], e piattaforme online, che fanno della comunicazione alla comunità dei loro utenti dei materiali prodotti da questi ultimi il loro principale modello di business[457]. In seno a queste ultime è poi individuata la categoria delle *Very Large Online Platform (VLOP)*, gravate di responsabilità ulteriori in ragione del ruolo dominante nel mercato dei servizi internet. La proposta identifica come *VLOP* le piattaforme che raggiungono, su base mensile, almeno 45 milioni di utenti locati sul suolo europeo[458].

Il *Digital service act* detta una serie di disposizioni che tutti gli intermediari coinvolti saranno tenuti ad applicare, e che, come già detto, riguardo al ruolo della tecnologia nell'attività di *contentmoderation* sembrano partire dai presupposti indicati poco sopra. L'apporto di sistemi avanzati, spesso basati sull'intelligenza artificiale, è visto, infatti, come irrunciabile, e la strada che la proposta normativa sceglie di percorrere è quella di una loro regolazione che valorizzi le possibilità di intervento umano. I prestatori di servizi internet sono tenuti, in primo luogo, a comunicare in modo chiaro e accessibile le regole che applicano nell'attività di moderazione, e queste dovranno essere applicate in modo diligente, obiettivo, proporzionato e rispettoso dei «diritti e degli interessi legittimi di tutte le parti coinvolte, compresi i diritti fondamentali applicabili dei destinatari del servizio» (art. 12)[459]. Pur in modo estremamente generico, dunque, per la prima volta

---

dell'anonimato, in *MediaLaws*, 12 gennaio 2021, https://bit.ly/3Qr6y9d (14 maggio 2020); G. ABALDO, *Una prospettiva di regolamentazione degli ISP attraverso il Digital Service Act*, in *MediaLaws*, 3 febbaio 2022, https://bit.ly/3xwgzcM (14 maggio 2022); A. NICITA, *Le piattaforme online tra moderazione e autoregolazione: verso il Digital Services Act*, in *MediaLaws*, 25 novembre 2020, https://bit.ly/3MS3iku (14 maggio 2022); GIOVANNI DE GREGORIO, *The Rise of Digital Constitutionalism in the European Union*, in *International Journal of Constitutional Law*, 19, 1, 2021, p. 41–70.

[455]Cfr. art. 3-4.

[456] Cfr. art. 5 e, per gli obblighi specifici dei prestatori di servizi di *hosting*, artt. 14-15.

[457] L'art. 2 par. 1 lett. h) della Proposta di Regolamento definisce "piattaforma online": «un prestatore di servizi di hosting che, su richiesta di un destinatario del servizio, memorizza e diffonde al pubblico informazioni, tranne qualora tale attività sia una funzione minore e puramente accessoria di un altro servizio e, per ragioni oggettive e tecniche, non possa essere utilizzata senza tale altro servizio e a condizione che l'integrazione di tale funzione nell'altro servizio non sia un mezzo per eludere l'applicabilità del presente regolamento». Per gli obblighi in capo alle piattaforme, cfr. artt. 16-24.

[458]Per gli oneri di cui sono gravate le piattaforme di grandi dimensioni, cfr. artt. 25 ss. L'art. 25 par. 1, appunto, specifica che «la presente sezione si applica alle piattaforme online che prestano i loro servizi a un numero medio mensile di destinatari attivi del servizio nell'Unione pari o superiore a 45 milioni».

[459]L'art. 12 (*condizioni generali*) recita: «I prestatori di servizi intermediari includono nelle loro condizioni generali informazioni sulle restrizioni che impongono in relazione all'uso dei loro servizi per quanto riguarda le informazioni

una norma di legge espliciterebbe, in caso di approvazione, il legame necessario che esiste tra termini di servizio delle piattaforme e diritti fondamentali, libertà d'espressione *in primis*. Tutti gli operatori, inoltre, sono gravati da precisi obblighi di trasparenza, consistenti nella pubblicazione, a cadenza almeno annuale, di un *report* dettagliato sull'attività di moderazione da essi condotta (art. 13)[460].

Accanto a queste obbligazioni comuni a tutte le categorie, la proposta di *Digital Service Act* propone una lunga serie di adempimenti, graduati proporzionalmente al livello di rischio e all'importanza degli operatori coinvolti, adottando un approccio già utilizzato, come già detto, per il GDPR e che ispira anche l'intera proposta di Regolamento in materia di IA. *Provider* di servizi di *hosting* e piattaforme online sono tenuti, nello specifico, a mettere in funzione un sistema di *notice and take down*, in modo da essere in grado di rimuovere in tempi rapidi i contenuti illeciti segnalati dagli utenti (un obbligo cui sono sottratti i puri intermediari, non conservando a lungo termine i contenuti degli utenti, e che, nella pratica, gran parte delle piattaforme già assolve spontaneamente)[461]. Sulle piattaforme online gravano una serie di ulteriori oneri, in ragione dei rischi connessi al loro modello di *business*, incentrato sulla diffusione al pubblico dei contenuti degli utenti, e del possibile impatto sul discorso pubblico e la libertà di manifestazione del pensiero della loro attività. Si tratta di un insieme di adempimenti volto a creare un contesto *online* il più possibile al riparo dalla diffusione di contenuti illeciti o malevoli e in grado di garantire, al contempo, un livello di pluralismo e libertà d'espressione consono agli standard delle società democratiche. Tra gli obblighi più significativi imposti alle piattaforme online vi è, in particolare, la creazione di effettivi meccanismi con cui gli utenti possano chiedere una revisione motivata di eventuali decisioni di moderazione[462], il rispetto di stringenti requisiti di trasparenza sui finanziatori

---

fornite dai destinatari del servizio. Tali informazioni riguardano tra l'altro le politiche, le procedure, le misure e gli strumenti utilizzati ai fini della moderazione dei contenuti, compresi il processo decisionale algoritmico e la verifica umana. Sono redatte in un linguaggio chiaro e privo di ambiguità e sono disponibili al pubblico in un formato facilmente accessibile.

I prestatori di servizi intermediari agiscono in modo diligente, obiettivo e proporzionato nell'applicare e far rispettare le restrizioni di cui al paragrafo 1, tenendo debitamente conto dei diritti e degli interessi legittimi di tutte le parti coinvolte, compresi i diritti fondamentali applicabili dei destinatari del servizio sanciti dalla Carta.».

[460] A stabilirlo è la prima parte dell'art. 13 par. 1: «I prestatori di servizi intermediari pubblicano, almeno una volta all'anno, relazioni chiare, facilmente comprensibili e dettagliate sulle attività di moderazione dei contenuti svolte durante il periodo di riferimento».

[461] Lo prevede l'art. 14 della Proposta di Regolamento, il cui par. 1 recita: «I prestatori di servizi di hosting predispongono meccanismi per consentire a qualsiasi persona o ente di notificare loro la presenza nel loro servizio di informazioni specifiche che tale persona o ente ritiene costituiscano contenuti illegali. Tali meccanismi sono di facile accesso e uso e consentono la presentazione di notifiche esclusivamente per via elettronica». I paragrafi successivi specificano i requisiti che tali sistemi di *notice and take down* sono tenuti a soddisfare.

[462] È l'articolo 17 a stabilire presupposti, modalità e termini del reclamo: «Le piattaforme online forniscono ai destinatari del servizio, per un periodo di almeno sei mesi dalla decisione di cui al presente paragrafo, l'accesso a un sistema interno di gestione dei reclami efficace, che consenta di presentare per via elettronica e gratuitamente reclami contro le seguenti decisioni adottate dalla piattaforma online a motivo del fatto che le informazioni fornite dai destinatari costituiscono contenuti illegali o sono incompatibili con le sue condizioni generali:a)le decisioni di rimuovere le

della pubblicità online[463], e la predispozione, al fine di limitare l'utilizzo impreciso, strumentale o illecito dei meccanismi di *notice and take down*, di gruppi di *trusted flaggers*, le cui segnalazioni di contenuti illeciti dovranno essere considerate prioritarie e particolarmente affidabili[464]. La proposta di Regolamente prevede, inoltre, in modo particolarmente penetrante e innovativo, la predisposizione di organi di risoluzione stragiudiziale delle controversie cui gli utenti potranno rivolgersi per il riesame di decisioni attinenti all'attività di *content moderation*, tenuti a soddisfare determinate garanzie di indipendenza dalle piattaforme, verificate da autorità pubbliche degli Stati membri[465]. L'eventuale ricorso a tali organi non precluderà, in ogni caso, la possibilità di rivolgersi all'Autorità giudiziaria. Sulle *Very Large Online Platform*, infine, grava la responsabilità di condurre una stringente, complessa e dettagliata attività di valutazione, gestione e mitigazione dei rischi sistemici connessi al loro modello di *business*, anche attraverso procedure di *auditing* esterno. Tali operatori, inoltre, sono gravati di oneri di trasparenza supplementari riguardo alla pubblicità associata ai contenuti[466] e sono tenuti a comunicare in modo chiaro e accessibile agli utenti le modalità di funzionamento dei sistemi di raccomandazione da essi eventualmente utilizzati[467]. Da

---

informazioni o disabilitare l'accesso alle stesse;b)le decisioni di sospendere o cessare in tutto o in parte la prestazione del servizio ai destinatari;c)le decisioni di sospendere o cessare l'account dei destinatari».

[463]L'obbligo è previsto dall'art. 24: «Le piattaforme online che visualizzano pubblicità sulle loro interfacce online provvedono affinché i destinatari del servizio siano in grado di identificare in modo chiaro e non ambiguo e in tempo reale, per ogni singolo messaggio pubblicitario mostrato a ogni singolo destinatario:a) la natura pubblicitaria delle informazioni visualizzate;b) la persona fisica o giuridica per conto della quale viene visualizzata la pubblicità;c) informazioni rilevanti sui principali parametri utilizzati per determinare il destinatario al quale viene mostrata la pubblicità».

[464]Cfr. l'art. 19 della Proposta. Le caratteristiche principali che i *trusted flaggers* devono soddisfare sono elencate al par. 2: «La qualifica di segnalatore attendibile a norma del presente regolamento viene riconosciuta, su richiesta di qualunque ente, dal coordinatore dei servizi digitali dello Stato membro in cui è stabilito il richiedente, a condizione che quest'ultimo abbia dimostrato di soddisfare tutte le condizioni seguenti: a) dispone di capacità e competenze particolari ai fini dell'individuazione, dell'identificazione e della notifica di contenuti illegali; b) rappresenta interessi collettivi ed è indipendente da qualsiasi piattaforma online;c)svolge le proprie attività al fine di presentare le notifiche in modo tempestivo, diligente e obiettivo».

[465]Tali organi di risoluzione stragiudiziale delle controversie saranno incaricati, in primo luogo, della revisione delle decisioni dei reclami presentati ai sensi dell'art. 17, e della valutazione dei casi in cui la procedura di reclamo interno risulti non esperibile. La disciplina di tali nuovi organi, e delle modalità della loro certificazione, è dettata dall'art. 18 della Proposta.

[466] L'art. 30 del *Digital Services Act*, in particolare, prevede che le piattaforme di grandi dimensioni compilino e rendano accessibile al pubblico un registro in cui, per ogni iniziativa in materia di pubblicità, sia annotato: «a) il contenuto della pubblicità; b) la persona fisica o giuridica per conto della quale viene visualizzata la pubblicità; c) il periodo durante il quale è stata visualizzata la pubblicità; d) un'indicazione volta a precisare se la pubblicità fosse destinata ad essere mostrata a uno o più gruppi specifici di destinatari del servizio e, in tal caso, i principali parametri utilizzati a tal fine; e) il numero totale di destinatari del servizio raggiunti e, ove opportuno, i dati aggregati relativi al gruppo o ai gruppi di destinatari ai quali la pubblicità era specificamente destinata».

[467]Lo provede l'art. 29 della Proposta, unitamente a forme inedite di coinvolgimento dell'utente nell'attivazione e nel funzionamento dei sistemi di raccomandazione: «Le piattaforme online di dimensioni molto grandi che si avvalgono di sistemi di raccomandazione specificano nelle loro condizioni generali, in modo chiaro, accessibile e facilmente comprensibile, i principali parametri utilizzati nei loro sistemi di raccomandazione, nonché qualunque opzione che possano avere messo a disposizione dei destinatari del servizio per consentire loro di modificare o influenzare tali parametri principali, compresa almeno un'opzione non basata sulla profilazione ai sensi dell'articolo 4, punto 4), del regolamento (UE) 2016/679.

Qualora siano disponibili più opzioni a norma del paragrafo 1, le piattaforme online di dimensioni molto grandi mettono a disposizione una funzionalità facilmente accessibile sulla loro interfaccia online che consenta ai destinatari del

ultimo, in caso di approvazione della proposta le grandi piattaforme dovranno condividere con le autorità degli stati membri e con la comunità scientifica i dati necessari a monitare e valutare la conformità della loro condotta alla normativa in esame e a condurre ricerche finalizzate allo sviluppo di nuove strategie di gestione dei rischi sistemici associati al loro funzionamento[468].

Le ipotesi e prospettive *de iure condendo* esaminate fin qui, e le cautele imposte alle piattaforme da esse previste, di certo darebbero un contributo importante alla risoluzione di molti degli interrogativi sollevati dall'attività di *content moderation* e in precedenza estensivamente commentati. Non è, però, ragionevole aspettarsi che esse siano sufficienti a risolvere completamente diversi di tali problemi. Infatti, come più volte ripetuto, una delle ragioni essenziali dell'attuale struttura della *content moderation* sta nella dimensione delle piattaforme, da cui deriva una quantità eccezionale di contenuti da gestire e un'estrema diversità in seno a questi ultimi in termini di natura, tematiche, lingua e cultura di riferimento. Gli accorgimenti teorizzati fin qui sono di certo idonei ad aumentare le garanzie per gli utenti e a migliorare precisione, granularità ed efficacia dell'attività di moderazione dei contenuti su larga scala. Un numero – elevato in valore assoluto, anche se in proporzione estremamente ridotto – di casi critici continuerebbe comunque ad esistere, poiché sono le dimensioni stesse del problema a renderlo intrattabile. Da un lato,alcunicontenuti, la cui diffusione potrebbe essere percepita come molto importante dagli utenti in questione,continuerebbero a venire moderati in base a motivazioni opinabili, o veri e propri errori del sistema di moderazione, nella sua componente automatizzata o umana. Contro tali decisioni non rimarrebbero che le procedure di riesame interne alle piattaforme – in teoria molto rapide, ma esposte ovviamente a disfunzioni e alla possibilità che si presentino carichi di lavoro eccessivi – o, in ultima analisi, il ricorso giurisdizionale, con ogni conseguenza, economica e pratica, che spesso ne deriva. Dall'altro lato, materiali potenzialmente idonei a generare un pericolo concreto per l'ordine pubblico, come disinformazione su temi particolarmente delicati o discorsi d'odio, potrebbero sfuggire anche al più perfetto dei sistemi di moderazione e infiammare il dibattito pubblico.Specialmente in riferimento a questi ultimi rischi, pare meritevole di interesse un'ipotesi avanzata in letteratura, che, al fine di mitigare le conseguenze di eventuali disfunzioni dell'attività di moderazione, propone di imporre alle piattaforme con gli strumenti del diritto una modifica alla struttura dei loro servizi che si differenzia, rispetto a tutte quelle ipotizzate finora, per essere indifferente alla natura dei contenuti di volta in volta in esame. Un approccio *content-blind* che

servizio di selezionare e modificare in qualsiasi momento l'opzione da essi preferita per ciascuno dei sistemi di raccomandazione che determina l'ordine relativo delle informazioni loro presentate».

[468]Lo prevede l'art. 31 della Proposta, specificando che, per avere accesso ai dati, i ricercatori «devono essere affiliati a istituzioni accademiche, essere indipendenti da interessi commerciali, disporre di comprovate competenze nei settori connessi ai rischi esaminati o alle relative metodologie di ricerca e devono assumere l'impegno ed essere in grado di rispettare gli specifici requisiti di sicurezza e riservatezza dei dati corrispondenti a ciascuna richiesta».

avrebbe il pregio di evitare che la soluzione delle problematiche conesse alla *content moderation* dipenda dalla delega, totale o parziale, a piattaforme private di delicate decisioni attinenti alla libertà di manifestazione del pensiero e spesso connotate da un inevitabile contenuto politico. Si tratta, molto semplicemente, di limitare la diffusione di ogni contenuto diffuso sulle piattaforme, a prescindere dalla natura di quest'ultimo, imponendo un limite alla possibilità di condivisione nello spazio e nel tempo (impedendo, ad esempio, che un *post* possa essere rilanciato più di mille volte in un'ora, da uno stesso o da diversi profili)[469]. Questo farebbe sì che qualunque contenuto, per quanto disturbante, non possa venire visualizzato, in tempi brevissimi, da milioni di persone, e consentirebbe di intervenire, eventualmente, con l'eventuale attività di moderazione *ex post* che risultasse necessaria. La mitigazione delle possibili conseguenze derivanti dalla diffusione di contenuti malevoli – destinati a raggiungere, in ogni caso, un numero di utenti più limitato di quello attuale – permetterebbe, probabilmente, di adottare criteri di moderazione *ex ante* meno stringenti, e le piattaforme, in generale, avvertirebbero in misura minore la tentazione di ricorrere a forme di censura preventiva al fine di evitare possibili problemi. La limitazione alla possibilità di condivisione e rilancio dei contenuti diffusi da un utente potrebbe avere una durata limitata nel tempo (ad esempio, nei termini di alcune ore) al solo fine di permettere ai primi utenti che li visionano di richiedere, se lo ritengono necessario, l'intervento della piattaforma, o risultare permanente, ridisegnando, così, l'intera struttura di molti servizi dell'Internet 2.0. Nonostante quest'ultima soluzione potrebbe apparire, a un primo sguardo, apertamente censoria, limitando la capacità di azione di ogni utente di tali servizi, preme evidenziare come essa, in realtà, metterebbe fine all'attuale predominio di determinati operatori che, con investimenti spesso ingenti, sfruttano i meccanismi della viralità per occupare uno spazio sulle piattaforme sproporzionato rispetto alla loro reale importanza nel dibattito pubblico. L'ovvia conseguenza indiretta è la penalizzazione dei contenuti di persone comuni, o di media che non adottano le stesse strategie. Si tratta in molti casi degli account di riferimento di partiti politici, associazioni o movimenti d'opinione di ogni genere, ma anche di soggetti che, spesso protetti dal relativo anonimato garantito da internet, diffondono contenuti controversi, talvolta per conto di mandanti e finanziatori non chiari[470]. Limitare la

[469] La proposta di mettere in atto accorgimenti di questo genere è di S. QUINTARELLI, *Capitalismo immateriale. Le tecnologie digitali e il nuovo conflitto sociale*, Torino, 2019. Per uno spunto simili, inoltre, sia concesso, inoltre, rinviare a L. RINALDI, *Le piattaforme tra diritto pubblico e diritto privato cit.*, p. 226 ss.

[470] Cfr. ancoraN. GRINBERG ET AL., *Fake News on Twitter during the 2016 U.S. Presidential Election cit.*; M. DEL VICARIO ET AL., *The Spreading of Misinformation Online cit.*; E. FERRARA ET AL., *The Rise of Social Bots cit.*; M. WOO, *How Online Misinformation Spreads cit.* Riguardo al finanziamento e alla monetizzazione dei contenuti malevoli diffuse su internet, invece, cfr. G. ROSIE, *Google and advertising: digital capitalism in the context of post-fordism, the reification of language, and the rise of fake news*, in *Palgrave communications*, 3, 1, p. 1–19; J. A. BRAUN, J. L. EKLUND, *Fake News, Real Money: Ad Tech Platforms, Profit-Driven Hoaxes, and the Business of Journalism*, in *Digital Journalism*, 7, 1, p. 1–21; D. TAMIBINI, *How advertising fuels fake news*, in *LSE Media policy blog*, 2017, https://bit.ly/3mT3uoM (14 maggio 2022).

possibilità di ogni contenuto di diventare virale, imponendo un limite ragionevole alla sua velocità di diffusione, potrebbe, quindi, in realtà rendere il webun luogo più democratico e le piattaforme più simili allo spazio in cui ogni individuo acquisiva una possibilità concreta di esprimersi e di rivolgersi a un pubblico relativamente numeroso che promettevano di essere ai loro inizi. L'idea, quindi, non pare da scartare a prescindere e, *mutatis mutandis*, ricorda alla lontana, almeno nel presupposto generale che l'attenzione e il tempo degli utenti non siano una risorsa infinita, e che le modalità di occupazione di questi ultimi debbano, quindi, essere regolamentate, i meccanismi di parcellizzazione degli spazi telesivi finalizzati a una loro gestione pluralistica in vigore in molto ordinamenti[471].

Da ultimo, deve segnalarsi un'ipotesi di lavoro attinente alla struttura della regolazione delle piattaforme e della *content moderation* e non al suo contenuto, proposta anche da autorevoli voci della dottrina italiana: quella di porre, a controllo dell'attività di moderazione, un'autorità indipendente[472]. I vantaggi di tale organo, da immaginareoperante al livello nazionale e coordinato da un'autorità capofila a livello europeo, sul modello di quanto avviene in materia di protezione dei dati personali, sarebbero, nell'ottica dei proponenti, molteplici. Lo schema dell'autorità indipendente permetterebbe, infatti, di emanare linee guida, pareri e norme tecniche anche molto dettagliate con la necessaria rapidità e flessibilità, superando, al contempo, l'attuale situazione di auto e co-relgolazione che si trasforma, nei fatti, in una delega pressoché in bianco agli operatori del settore. L'istituzione di un'apposita autorità permetterebbe, inoltre, di formalizzare il coinvolgimento di esperti di primo piano in un organo afferente ai pubblici poteri, conservando allo stesso tempo appropriate garanzie di indipendenza, nel tentativo di limitare sul nascere la comparsa di organi ibridi con natura e funzioni poco chiare creati dalle piattaforme, come il visto *Oversight Board* di *Facebook*. Non meno importante, la potestà sanzionatoria di cui tale organo sarebbe

[471] Limitandoci al caso italiano, è noto che la regolamentazione dei servizi radiotelevisivi al fine di tutelare il pluralismo e la libertà dell'informazione ha avuto una storia particolarmente travagliata, con un susseguirsi di normative spesso oggetto di sentenze della Corte costituzionale. Il passaggio al digitale terrestre, con il conseguente moltiplicarsi degli operatori televisi pubblici e privati, ha attenuato gli storici problemi di oligopolio. La materia è oggi disciplinata principalmente dal D. Lgs.n. 208 del 8 novembre 2021, *Attuazione della direttiva (UE) 2018/1808 del Parlamento europeo e del Consiglio, del 14 novembre 2018, recante modifica della direttiva 2010/13/UE, relativa al coordinamento di determinate disposizioni legislative, regolamentari e amministrative degli Stati membri, concernente il testo unico per la fornitura di servizi di media audiovisivi in considerazione dell'evoluzione delle realtà del mercato*(c.d. Testo Unico dei servizi media audiovisivi) e, per quanto riguarda la parità d'accesso al mezzo televisivo con fini di comunicazione politica (c.d. *par condicio*) dalla L. n. 28 del 22 febbraio 2000, *Disposizioni per la parità di accesso ai mezzi di informazione durante le campagne elettorali e referendarie e per la comunicazione politica*. In letteratura si vedano P. CARETTI, *La disciplina della radiotelevisione tra diritto interno e diritto comunitario*, in F. ANGOTTI, G. PELOSI, *Il telefono e dintorni: una selezione di eventi, contributi ed immagini dalle celebrazioni per il bicentenario della nascita di Antonio Meucci*, 2011, p. 125-128; A. VALASTRO, *Principi comuni a livello europeo in materia di propaganda elettorale televisiva*, in *Quaderni costituzionali*, 1, 1997, p. 109-130; F. CARDARELLI, V. ZENO-ZENCOVICH, *Il diritto delle telecomunicazioni: principi, normativa, giurisprudenza*, Bari, 1997.

[472] La proposta è di G. PITRUZZELLA, *La libertà d'informazione nell'era di Internet*, in O. POLLICINO, G. PITRUZZELLA, S. QUINTARELLI, *Parole e Potere cit.*; la riprende anche S. FOÀ, *Pubblici poteri e contrasto alle fake news. Versol'effettività dei diritti aletici?*, in *Federalismi.it*, 11, 2020,258-259.

potenzialmente dotato potrebbe dare un contributo decisivo al corretto funzionamento del sistema. La proposta, quindi, presenta indubbi punti a suo favore. Ciò nonostante, non può non sottolinearsi come vada trattata con particolare cautela, posto che l'istituzione di autorità non direttamente collegate ai circuiti elettivi e maggioritari, incaricate di vigilare sull'esercizio della libertà d'espressione, non può non ricordare le modalità di censura attuate dai regimi non democratici. Diverse voci, in primo luogo in Italia, hanno sollevato perplessità in proposito, in particolare con riferimento a quegli ambiti in cui la moderazione di contenuti è caratterizzata da maggiori ambiguità, come la disinformazione[473]. Deve essere evidenziato, d'altro canto, il valore delsistema di garanzie in cui un organo di tal genere andrebbe a inserirsi negli ordinamenti democratici contemporanei, compreso il nostro, che rende il paragone non particolarmente calzante. Il buon funzionamento delle autorità indipendenti create negli ultimi decenni in diversi ordinamenti e a livello europeopare, inoltre, un precedente tranquillizzante, in grado di mitigare di molto queste preoccupazioni[474]. Non si può non sottolineare, infine, come l'attuale situazione, in cui gran parte dei poteri censori sono spesso in mani quasi solo private, non appaia di certo connotata da maggor democraticità rispetto a quella che si disegnerebbe con la creazione dell'organo in esame. Anche l'ipotesi dell'istituzione di autorità amministrative indipendenti in materia di *contentmoderation*, dunque, una volta superate le perplessità iniziali, ampiamente infondate se si guarda con attenzione al concreto contesto di riferimento, non pare da scartare a prescindere, tra le possibilità per una regolazione futura efficace ed equilibrata.

[473] Cfr. in particolare N. ZANON, *Fake news e diffusione dei social media: abbiamo bisogno di un'"Autorità Pubblica della Verità"?*, in *Media Laws*, 1/2018, 15 ss. e, in riferimento a una vicenda puntuale, ma con considerazioni che si estendono all'ipotesi generale in esame, A. PRUITI CIARELLO, *Oggi la task force anti fake news. Domani? L'opinione della Fondazione Einaudi*, in *formiche.net*, https://bit.ly/2WGSEEU (15 maggio 2022); A. RINALDIS, *Una task force contro le false notizie? Il monopolio della verità è l'anticamera del totalitarismo*, in *Globalist.it*, 9 aprile 2020,https://bit.ly/2yvN6VY (15 maggio 2022).

[474] Tra i moltissimi contributi sul tema si rimanda, in via generale e da prospettive estremamente differenti, a M. CLARICH, *Autorità Indipendenti: Bilancio e Prospettive Di Un Modello*, Bologna, 2005; S. CASSESE, *Chi ha paura delle autorità indipendenti?*, in *Mercato Concorrenza Regole*, 3, 1999, p. 471–474, https://doi.org/10.1434/78; A. LA SPINA, S. CAVATORTO, *Le Autorità Indipendenti* , Bologna, 2008, p. 573 ss; M. LIBERTINI, *Autorità indipendenti, mercati e regole*, in *Rivista italiana per le scienze giuridiche*, 1, 2010, 63 ss.

Nuove sfide per "vecchi diritti". Intelligenza artificiale, discriminazione algoritmica e principio di eguaglianza

1. L'ascesa dell'intelligenza artificiale nei processi decisionali: le possibili discriminazioni e il ruolo del principio di eguaglianza

È intuitivo come il rischio di discriminazioni sia presente in ogni decisione o valutazione riguardante l'individuo presa da chi esercita su di esso un potere, formale o informale[475]. Il discorso, in realtà, può estendersi alla persona giuridica, e a ogni altro ente o centro d'imputazione di interessi giuridici. Dal punto di vista del possibile impatto negativo della decisione discriminatoria, la natura pubblica o privata del soggetto che se ne rende responsabile non fa differenza (dal punto di vista del diritto, la differenza spesso è netta, come vedremo). La presenza crescente, nei meccanismi decisionali, di sistemi automatizzati, nei quali l'intelligenza artificiale ha spesso un ruolo preponderante, va inquadrata in questo contesto[476].

È un fatto storico che la sempre maggiore complessità delle organizzazioni, pubbliche e private, abbia reso più stratificati, condivisi e laboriosi i meccanismi decisionali[477]. Non sono molti i casi in cui, nelle società industriali contemporanee, decisioni di rilievo sull'individuo siano il frutto della valutazione di uno o pochi soggetti (il sistema giurisdizionale, da questo punto di vista, rappresenta

[475] J. M. PEIRÓ, J. L. MELIÁ, *Formal and informal interpersonal power in organisations: testing a bifactorial model of power in role-sets*, in *Applied psicology*, 52, 1, 2003, p. 14-35; M. R. FERRARESE, *Privatizzazioni, poteri invisibili e infrastrutture giuridiche globali*, in *Diritto pubblico*, 3, 2021, p. 871-892; M. BETSU, *Poteri pubblici e poteri privati nel mondo digitale*, in *Rivista "Gruppo di Pisa"*, 2, 2021, p. 166-191; D.M. LAWRENCE, *Private exercise of governmental power*, in *Indiana Law Journal*, 61, 1986, p. 647-696.

[476] Sulle implicazioni del coinvolgimento dell'intelligenza artificial nei meccanismi decisionali si vedano, tra i moltissimi contributi sul tema e da prospettive differenti: Y. DUAN, J. S. EDWARDS, Y. K. DWIVEDI, *Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda*, in *International Journal of Information Management*, 48, 2019, p. 63–71; G. PHILLIPS-WREN, L. JAIN, *Artificial Intelligence for Decision Making*, in *Knowledge-Based Intelligent Information and Engineering Systems*, Berlino-Heidelberg, 2006, p. 531–536; J. C. POMEROL, *Artificial Intelligence and Human Decision Making*, in *European Journal of Operational Research* 99, 1, 1997, p. 3–25; K. DEAR, *Artificial Intelligence and Decision-Making*, in *The RUSI Journal*, 164, 5–6, p. 18–25; T. J. LOFTUS ET AL., *Artificial Intelligence and Surgical Decision-Making*, in *JAMA Surgery*, 155, 2, p. 148 ss.; Y. R. SHRESTHA, S. M. BEN-MENAHEM,G. VON KROGH, *Organizational Decision-Making Structures in the Age of Artificial Intelligence*, in*California Management Review*, 61, 4, 2019, p. 66–83; M. H. JARRAHI, *Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making*, in *Business Horizons*, 61, 4, 2018, p. 577–586; K. DE FINE LICHT, J. DE FINE LICHT, *Artificial Intelligence, Transparency, and Public Decision-Making: Why Explanations Are Key When Trying to Produce Perceived Legitimacy*, in *AI& SOCIETY*, 35, 4, 2020, p. 917–926; T. ARAUJO, N. HELBERGER, S. KRUIKEMEIER, C. H. DE VREESE, *In AI We Trust? Perceptions about Automated Decision-Making by Artificial Intelligence*, in *AI& SOCIETY*, 35, 3, 2020, p. 611–623; M. LUCIANI, *La decisionegiudiziariarobotica*, in *Rivista AIC*, 3, 2018, p. 872-893; D. D. LUXTON, *Should Watson Be Consulted for a Second Opinion?*, in*AMA Journal of Ethics*, 2, 2019, p. 131-138; T. SOURDIN, *Judge v. Robot? Artificial Intelligence and Judicialdecision-making*, in *UNSW Law Journal*, 4, 2018, p. 1114-1133; A. SANTOSUOSSO, *Intelligenza Artificiale e Diritto : Perché Le Tecnologie Di IA Sono Una Grande Opportunità per Il Diritto* , Milano, 2020; A. GARAPON, J. LASSEGUE, *Justice Digitale: révolution graphique et rupture anthropologique*, Parigi, 2018.

[477] Cfr. ad esempio E. BRUCH, F. FEINBERG, *Decision-Making Processes in Social Contexts*, in *Annual Review of Sociology*, 43, 1, 2017, p. 207–227;M. CRISTOFARO, *Reducing Biases of Decision-Making Processes in Complex Organizations*, in*Management Research Review*, 40, 3, 2017, p. 270–291; S. J. MILLER, D. J. HICKSON, D. C. WILSON, *Decision-making in organizations*, in S. R. CLEGG, C. HARDY, W. R. NORD (A CURA DI), *Managing organizations. Currentissues*, 1999, p. 43-63; E. MARCHELLO, *I processi decisionali*, Milano, 2003.

una vistosa eccezione). L'imputazione delle scelte a un individuo o a un organo identifica coloro che del contenuto e degli effetti di esse si assumono la responsabilità, ma non è idonea a definire il reale meccanismo decisionale, spesso frutto dell'attività (istruttoria, di consulenza, di discussione e valutazione) di diverse decine di persone. In questa situazione, l'uso di algoritmi rappresenta, allo stesso tempo, un elemento di semplificazione e di ulteriore complessità. L'intelligenza artificiale può rendere più semplici e snelli i processi decisionali, per l'intuitiva ragione di automatizzare ponderazioni e valutazioni che sarebbero, altrimenti, svolte da esseri umani (fino alla soluzione estrema in cui l'intera decisione è affidata all'algoritmo). Parallelamente, la complessità della scelta aumenta, perché le tecnologie intelligenti inseriscono nel meccanismo decisionale un elemento - la valutazione algoritmica – la cui gestione per l'operatore umano è particolarmente complessa. La mole di dati che tipicamente l'algoritmo considera, infatti, è di gran lunga superiore alle capacità dell'essere umano (del resto, in caso contrario non vi sarebbero serie ragioni per ricorrervi) e ciò rende particolarmente difficile l'elaborazione critica delle indicazioni della macchina.

La possibilità di esiti discriminatori non diminuisce o aumenta per il solo fatto del coinvolgimento, nella decisione, di tecnologie avanzate, e in particolare di sistemi di intelligenza artificiale. A determinare gli effetti concreti da tale punto di vista sono le caratteristiche dell'algoritmo e l'utilizzo che ne viene fatto. Le prossime pagine saranno dedicate all'analisi dei modi più rilevanti in cui l'utilizzo di tecnologie intelligenti in processi decisionali, o comunque in senso lato valutativi, ha un impatto sulla possibilità di discriminazioni in tali processi e dunque, dal punto di vista del diritto costituzionale, sulle varie declinazioni del principio di eguaglianza[478]. Il filo conduttore dell'indagine, al pari che nei capitoli precedenti, rimane la relazione tra l'intelligenza artificiale e il catalogo dei diritti fondamentali riconosciuto nella nostra tradizione costituzionale e

[478] Tra i numerosissimi studi sul principio di eguaglianza possono indicarsi, senza animo di completezza, C. ESPOSITO, *Eguaglianza e giustizia nell'art. 3 della Costituzione*, in C. ESPOSITO, *La Costituzione italiana. Saggi*, Padova, 1954; L. PALADIN, *Eguaglianza (diritto cost.)*, in *Enciclopedia del diritto*, XIV, Milano, 1965, p. 519 ss.; L. PALADIN, *Il principio costituzionale d'eguaglianza*, Milano, 1965; A. S. AGRÒ, U. ROMAGNOLI, *Commento all'art. 3 cost.*, in G. BRANCA (A CURA DI), *Commentario della Costituzione italiana*, Bologna, 1976; P. BISCARETTI DI RUFFIA, *Uguaglianza (principio di)*, in *Novissimo digesto italiano*, XIX, Torino, 1982, p. 1088 ss;A. CERRI, *Uguaglianza (principio costituzionale di)*, in *Enciclopedia giuridica Treccani*, Roma, 1988; M. CARTABIA, T. VETTOR (A CURA DI), *Le ragioni dell'uguaglianza*, Milano, 2009; F. SORRENTINO, *Eguaglianza*, Torino, 2011;A. CELOTTO, *Le declinazioni dell'eguaglianza*, Napoli, 2011; D. FLORENZANO, *Il principio costituzionale di eguaglianza*, in F. CORTESE, D. BORGONOVO RE, D. FLORENZANO, *Diritti inviolabili, doveri di solidarietà e principio di eguaglianza*, 2015, p. 103 ss. Per quanto riguarda gli studi di diritto straniero, comparato, internazionale e sovranazionale cfr. *ex multis*I. CASTANGIA, G. BIAGIONI, *Il principio di non discriminazione nel diritto dell'Unione europea*, Napoli, 2011; M. LUCIANI,*I principi di eguaglianza e non discriminazione: una prospettiva di diritto comparato*, EPRS-Servizio Ricerca del Parlamento Europeo, 2020; G. P. DOLSO, *Il principio di non discriminazione nella giurisprudenza della Corte Europea dei Diritti dell'Uomo*, Napoli, 2013; T. F. BOTTS, *For equals only: race, equality and the equal protection clause*, Lanham, 2018; F. M. SOUCRAMANIEN, *Le principe d''égalité dans la jurisprudence du Conseil constitutionnel*, Marsiglia, 1999; J. H. WILKINSON, *The Supreme Court, the Equal Protection Clause, and the Three Faces of Constitutional Equality*, in*Virginia Law Review*, 61, 5, 1975, p. 945 ss.; R. B. GINZBURG, *Interpretations of the equal protection clause*, in *Harvard Journal of Law & Public Policy*, 9, 1, 1986, p. 41-46; C. F. J. DOEBBLER, *Principle of non-discrimination in international law*, Washington, 2007; T. LOENEN, P. R. RODRIGUES (A CURA DI), *Non-discrimination law: comparative perspectives*, Utrecht, 1999.

in quelle ad essa più affini. Il principio di eguaglianza, infatti, è il presupposto per l'effettivo godimento di ogni diritto individuale, rappresentando il primo, fondamentale argine ai possibili comportamenti arbitrari dell'autorità. I paragrafi che seguono tratteranno, innanzitutto, delle diverse modalità con cui l'utilizzo di tecnologie dotate di componenti di intelligenza artificiale nei processi decisionali possa portare a risultati discriminatori, e di come tali discriminazioni rischino, spesso, di passare inosservate, a causa di una diffusa percezione acritica di oggettività e neutralità della tecnologia.

2. La presunta oggettività della tecnologia e gli ostacoli nell'identificazione della discriminazione algoritmica

Sono ormai numerosi gli studi che dimostrano che il coinvolgimento di tecnologie digitali in una determinata attività è accompagnato dalla convinzione – in molti casi fondata - che ciò riduca le possibilità di errore umano e porti a risultati più precisi, rapidi e uniformi[479]. Si tratta, del resto, del ruolo che l'innovazione tecnologica ha prevalentemente rivestito nel corso della storia, e in modo accentuato a partire dalla rivoluzione industriale. Lo sviluppo di macchinari per l'automazione, totale o parziale, di un numero sempre maggiore di attività, infatti, ha velocizzato e standardizzato la produzione di un amplissimo volume di beni. Prodotti in precedenza frutto del lavoro manuale tipico dell'artigiano, e quindi, a loro modo, "unici", hanno cominciato ad essere disponibili a prezzi più bassi e in versioni pressochè sempre identiche l'una all'altra. Parallelamente, il numero di difetti di lavorazione è calato drasticamente, poiché il macchinario industriale non era soggetto al margine d'errore di un essere umano, né alle imperfezioni dovute a stanchezza, malattia e distrazioni in cui quest'ultimo potesse incorrere[480]. Lo sviluppo del personal computer è stato accompagnato dalle

[479] E. M. CAMPBELL, D. F. SITTIG, K. P. GUAPPONE, R. H. DYKSTRA, J. S. ASH, *Overdependence on technology: An unintended adverse consequence of computerized provider order entry*, in *AMIA Annual Symposium Proceedings*, 2007, p. 94– 98, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2710605/ (16 maggio 2022); V. ARNOLD, P. COLLIER, P. S. LEECH, S. G. SUTTON, *Impact of intelligent decision aids on expert and novice decision-makers' judgments*, in *Accounting & Finance*, 44, 1, 2004, p. 1– 26.; V. ARNOLD, S. G. SUTTON, *The theory of technology dominance: Understanding the impact of intelligent decision aids on decision maker's judgments*, in *Advances in Accounting Behavioral Research*, 1, 3, 1998 p. 175– 194; C. KEDING, P. MEISSNER, *Managerial Overreliance on AI-Augmented Decision-Making Processes: How the Use of AI-Based Advisory Systems Shapes Choice Behavior in R&D Investment Decisions*, in *Technological Forecasting and Social Change*, 171, 2021 https://www.sciencedirect.com/science/article/pii/S0040162521004029 (16 maggio 2022); M. ADAMS, *Over-reliance on technology may be at the expense of care*, in *Nursing standard*, 29,6, 2014, p. 32-33; E.GLIKSON, A. WILLIAMS WOOLLEY, *Human Trust in Artificial Intelligence: Review of Empirical Research*, in *Academy of Management Annals* 14, 2, 2020, p. 627–660; M. SCHEMMER, P. HEMMER, N. KÜHL, C. BENZ,G. SATZGER, *Should I Follow AI-based Advice? Measuring Appropriate Reliance in Human-AI Decision-Making*, in *Conference on Human Factors in Computing Systems 2022*, *Workshop on trust and reliance in AI-human teams (trAIt)*, New Orleans, 2022, J. WU, J. THORNE-LARGE, P. ZHANG, *Safety First: The risk of over-reliance on technology in navigation*, in*Journal of Transportation Safety & Security*, 2021, p. 1–28 https://bit.ly/3N5z65y (15maggio 2022).

[480] Cfr. in generale E. DE SIMONE, *Storia economica. Dalla rivoluzione industriale alla rivoluzione informatica*, Milano, 2014; J. DE VRIES, *The industrial revolution and the industrious revolution*, in *The journal of economic history*, 54, 2, 1994, p. 249-270; P. HUDSON, *The industrial revolution*, Londra, 2014, p. 166 ss.; G. CLARK, *The industrial revolution*, in P. AGHION, S. N. DURLAUF, *Handbook of economic growth*, II, Amsterdam, 2014, p. 217-262.

stessesensazioni diffuse, rendendo immensamente più precise, facili e meno soggette a possibili errori una grande varietà di attività di calcolo, progettazione, scrittura, disegno o archiviazione di informazioni e documenti [481]. Pare, dunque, fisiologico che questa percezione di efficienza, oggettività e neutralità esista anche verso le tecnologie riconducibili alla famiglia dell'intelligenza artificiale[482].

La situazione, in quest'ultimo caso, diventa, però, differente e molto più complessa. I sistemi di intelligenza artificiale impiegati per supportare – o svolgere totalmente – processi decisionali e valutativi si basano, come già detto, sull'analisi dei dati. A prescindere dalla tecnologia concretamente coinvolta (il macrosettore di riferimento, in ogni caso, è comunemente l'apprendimento automatico) l'output di questo genere di sistemi deriva dall'elaborazione di una grande mole di precedenti riguardanti casi simili a quello che viene loro sottoposto[483]. Si tratta, in ultima analisi, di un risultato frutto di una valutazione di tipo statistico, per quanto basata su una quantità di dati del tutto fuori dalla portata di un essere umano e, dunque, estremamente precisa[484]. Come tale, l'indicazione di un algoritmo rimane sempre perfezionabile e soggetta a un margine d'errore ineliminabile, per quanto ridotto [485]. In sintesi, gli strumenti di intelligenza artificialecomunemente utilizzati nei processi decisionali sono tecnologie qualitativamente diverse da quelle che hanno caratterizzato la rivoluzione industriale e, poi, quella digitale. La percezione di oggettività e infallibilità che queste ultime hanno generato era pienamente calzante e giustificata, ma dovrebbe essere almeno in parte abbandonataper quanto riguarda le tecnologie ora in esame. Essa, infatti, andrebbe sostituita con la diversa consapevolezza dell'elevata affidabilità di tali

[481] Cfr. J. FITZSIMMONS, *Information technology and the third industrial revolution*, in *The Electronic Library*, 12, 5, p. 295-297; M. CASTELLS, *The Rise of the Network Society - The Information Age: Economy, Society and Culture*, I, Oxford, 2000; C. COOPER, *Technology and development in the industrial devolution*, Londra, 2005.

[482] Non a caso, l'intelligenza artificiale è tra gli elementi principali di quella che è stata definita, in un'ideale continuità con le precedenti, "quarta rivoluzione industriale". Cfr. ad es. T. PHILBECK, N. DAVIS, *The fourth industrial devolution: shaping a new era*, in *Journal of International affairs*, 72, 1, 2018, p. 17-22; N. DAVIS, *What is the fourth industrial revolution?*, World Economic Forum, 16 gennaio 2016, https://bit.ly/2HgFmd2 (14 maggio 2022); K. SCHWAB, *The fourth industrial revolution*, New York, 2017.

[483] Cfr. D. E. O'LEARY, *Artificial Intelligence and Big Data*, in *IEEE Intelligent Systems*, 28, 2, 2013, p. 96-99; S. A. YABLONSKY, *Multidimensional data-driven artificial intelligence cit.*; H. R. VARIAN, *Beyond big data cit.*; S. BAROCAS, A. D. SELBST, *Big Data's Disparate Impact cit.*; Y.DUAN, J. S. EDWARDS, Y. K. DWIVEDI, *Artificial intelligence for decision making cit.*; G. PHILLIPS-WREN, L. JAIN, *Artificial Intelligence for Decision Making cit.*; J. C. POMEROL, *Artificial Intelligence and Human Decision Making cit.*; K. DEAR, *Artificial Intelligence and Decision-Making cit.*; G. PHILLIPS-WREN, L. JAIN, *Artificial Intelligence for Decision Making cit*; J. C. POMEROL, *Artificial Intelligence and Human Decision Making cit.*

[484] S. QUINTARELLI, F. COREA, F. FOSSA, A. LOREGGIA, S. SAPIENZA, *AI: Profili etici. Una prospettiva etica sull'intelligenza artificiale: principi, diritti e raccomandazioni*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2019, p. 195 ss. parlano dell'intelligenza artificiale applicata ai processi decisionali come «motore statistico che produce necessariamente risultati probabilistici».

[485] Cfr. ad es. C. Q. CHOI, *7 revealing ways AIs fail,* in *IEEE Spectrum*, 21 settembre 2021, https://spectrum.ieee.org/ai-failures#toggle-gdpr (14 maggio 2022); R. YAMPOLSKIY, *AI will fail, like everything else, eventually*, in *Mind matters news*, 14 luglio 2020, https://mindmatters.ai/2020/07/ai-will-fail-like-everything-else-eventually/ (14 maggio 2022); A. H. RUTKIN, *The tiny changes that can cause AI to fail*, BBC – Future Now, 11 aprile 2017, https://www.bbc.com/future/article/20170410-how-to-fool-artificial-intelligence (16 maggio 2022).

strumenti, a patto di tenere a mente l'esistenza, inevitabile, di un margine d'errore e di possibili inesattezze. Si tratta di un tema che renderà necessario, nei decenni futuri, un imponente sforzo di sensibilizzazione, la cui messa in atto, peraltro, appare complicata dalla presenzadi chiaretendenze di segno contrario nell'attuale panorama culturale. Diversi studi, ad esempio, hanno messo in luce come l'essere umano, affiancato dall'intelligenza artificiale in attività valutative o decisionali, sia particolarmente soggetto a quella che è stata definita *distorsione dell'automazione* o *effetto moutonnier* (efficacemente tradotto in italiano con l'espressione "effetto pecorone")[486]: la tendenza a fidarsi eccessivamente della macchina, diminuendo gradualmente la propria attenzione e la profondità della revisione critica delle indicazioni di quest'ultima. È intutivo come ciò possa portare, sul lungo periodo, alla diminuzione delle stesse conoscenze e competenze che sarebbero necessarie per prendere la decisione senza la macchina[487]. Non solo, dunque, gli strumenti di intelligenza artificiale generalmente coinvolti nei processi decisionali sono percepiti come più neutrali e oggettivi di quanto in realtà non siano, a causa di una percezione culturale della tecnologia sviluppata prima del loro avvento, ma l'essere umano tende addirittura ad affidarsi ad essi in modo acritico in misura maggiore a quanto generalmente avviene con le altre macchine.

3. Le diverse tipologie di *bias* algoritmico e le loro conseguenze discriminatorie

Vi sono varie ragioni per cui un sistema informatico può dar adito a discriminazioni. La mancanza di neutralità da parte di un algoritmo è spesso indicata con la parola *bias*, che, nel linguaggio comune, ha ormai assunto una connotazione quasi esclusivamente negativa. In realtà, il termine in origine esprimeva ogni deviazione da un determinato standard di funzionamento[488]. Riguardo al tema della discriminazione algoritmica, con *bias* viene indicata ogni ragione che porta l'algoritmo a un risultato diverso da quello cui dovrebbe condurre il suo normale funzionamento[489]. Si tratta di un

[486] L'utilizzo dell'espressione in riferimento all'intelligenza artificiale è di A. GARAPON, J. LASSEGUE, *Justice Digitale cit.* Si vedano anche E. FRONZA, *"Code is Law". Note a margine del volume di Antoine Garapon e Jean Lassègue, Justice Digitale. Révolutiongraphique et ruptureanthropologique, Puf, Paris*, in *Diritto Penale Contemporaneo*, 11 dicembre 2018, https://bit.ly/3HsXBbA (14 maggio 2022); J. DE CODT, *Justice et algorithme: danger pour le procèséquitable et la démocratie?*, in *Revuetrimestrielledes droits de l'homme*, 117, 1, 2019, p. 3-11; B. MARCHETTI, *La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica*, in *BioLaw Journal - Rivista di BioDiritto*, 2, 2021 p. 367-385; C. CASONATO, *L'intelligenza artificiale e il diritto pubblico comparato ed europeo*, in *DPCE Online*, 51, 1, 2022, http://193.205.23.57/index.php/dpceonline/article/view/1566 (15 maggio 2022); C. CASONATO, *Giustizia e intelligenza artificiale: considerazioni introduttive*, in *BioLaw Journal - Rivista di BioDiritto*, 2, 2021 p.359-365.

[487] Cfr. ad es. J. LU, *Will Medical Technology Deskill Doctors?*, in *International Education Studies*, 9, 7, 2016, p. 130–134; J. LEVY, A. JOTKOWITZ, I. CHOWERS, *Deskilling in Ophthalmology Is the Inevitable Controllable?*, in *Eye*, 33, 3, 2019, p. 347–348; S. DE PAOLI, *Automatic-Play and Player Deskilling*, in *MMORPGs,Game Studies*, 13, 1, 2013; E. SINAGRA, F. ROSSI, D. RAIMONDO, *Use of Artificial Intelligence in Endoscopic Training: Is Deskilling a Real Fear?*, in *Gastroenterology* 160, 6, 2021, p. 2212 ss.

[488] L'etimologia del termine, infatti, è ricondotta al francese *biais*, significante semplicemente "orientamento, piega, inclinazione", v. *Bias (n.)*, in *Online Etymology Dictionary*, https://www.etymonline.com/word/bias (15 maggio 2022).

[489] B. FRIEDMAN, H. NISSENBAUM, *Bias in computer systems*, in *ACM Transactions on Information Systems*, 3, 1996, doi.org/10.1145/230538.230561; R. DOBBES. DEAN, T. GILBERT, N. KOHLI, *A Broader View on Bias in Automated*

fenomeno molto variegato, e sono state avanzate diverse proposte di suddivisione dei *bias* algoritmici in categorie differenti, a seconda delle loro caratteristiche[490]. Semplificando, possiamo identificare quattro principali tipologie di *bias*. In primo luogo, il caso in cui i risultati discriminatori dell'algoritmo derivano da *bias* nei *training data*, in cui talune variabili risultano sotto o sovrarappresentate, e questa imperfezione si riflette, fatalmente, negli *outcome* del sistema[491]. In secondo luogo, il *bias* può essere presente nella stessa struttura dell'algoritmo, ad esempio al fine di ridurre il peso di alcuni dati di *input* anomali o inutili ai fini delle funzioni per cui è impiegato. Si tratta di strategie impiegate nello sviluppo dei sistemi per migliorare la qualità dei loro risultati, e il cui effetto normalmente è positivo. Ciò non toglie che incidano sulla "neutralità" dell'algoritmo, e possano portare a risultati discriminatori – posto che implicano di assegnare gradi d'importanza differenti a determinate variabili - non immaginati al momento della progettazione[492]. In terzo luogo, il *bias* può sorgere non dai dati d'addestramento, né dall'algoritmo, ma dall'utilizzo di un algoritmo di per sé "neutrale" in un contesto di riferimento diverso da quello per cui è stato concepito[493]. L'esempio che può farsi, particolarmente pregnante, è quello di un sistema di guida autonoma addestrato con dati relativi al traffico stradale in un paese dell'Europa continentale, ma utilizzato sulle strade della Gran Bretagna[494]. Sono intuibili gli effetti che potrebbero derivarne, posto che il veicolo non è stato addestrato per la guida a sinistra.

Ciascuna di queste prime tre categorie di *bias* può essere generata da semplice errore umano (è il caso della formazione di un *dataset* di addestramento non accurato), da intenzioni malevole e volontarie dello sviluppatore o può rappresentare, semplicemente, la traduzione in termini algoritmici di un pregiudizio presente in modo inconsapevole in chi progetta e addestra l'algoritmo. In quest'ultimo caso, il *dataset* trascura alcune variabili perché sono gli stessi esseri umani che lo predispongono che, inavvertitamente, ne sminusicono l'importanza e il reale tasso di diffusione; l'algoritmo è progettato in modo discriminatorio perché riproduce un'erronea percezione della

Decision-Making: Reflecting on Epistemology and Dynamics, 2018, arXiv:1807.00553 (14 maggio 2022); J. SILBERG, J. MANYIKA., *Notes from the AI frontier: Tackling bias in AI (and in humans)*, in *McKinsey Global Institute*, 2019, https://mck.com/3ih2l6L.

[490] Cfr. ad es. D. DANKS, A. J. LONDON, *Algorithmic bias in autonomous systems*, in*Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*, 17, 2017, p. 4691-4697;S. SILVA, M. KENNEY, *Algorithms, Platforms, and Ethnic Bias*, in *Communications of the ACM*, 62, 11,p. 37–39; S. SILVA, M. KENNEY, *Algorithms, Platforms, and Ethnic Bias: An Integrative Essay*, in *Phylon*, 55, 1-2, 2018, p. 9–37; B. FRIEDMAN, H. NISSENBAUM, *Bias in computer systems cit*.

[491] B. MAC NAMEE, P. CUNNINGHAM, S. BYRNE, O.I. CORRIGAN, *The problem of bias in training data in regression problems in medical decision support*, in *Artificial Intelligence in Medicine*, 24, 1,2002, p. 51-70; D. DANKS, A. J. LONDON, *Algorithmic bias in autonomous systems*, p. 2-3.

[492] S. HOOKER, *Moving beyond "algorithmic bias is a data problem"*, in *Patterns*,2,4,2021,https://doi.org/10.1016/j.patter.2021.100241; A. SŁOWIK, L. BOTTOU, *Algorithmic Bias and Data Bias: Understanding the Relation between Distributionally Robust Optimization and Data Curation*, 2021,https://arxiv.org/abs/2106.09467 (16 maggio 2022).

[493] Cfr. S. SILVA, M. KENNEY, *Algorithms, platforms, and ethnic bias*e*Algorithms, platforms, and ethnic bias: an integrative essay cit.*

[494] Portano questo esempio D. DANKS, A. J. LONDON, *Algorithmicbias in autonomoussystems*, p. 4.

realtà di chi lo sviluppa; oppure il sistema è utilizzato inavvertitamente in un contesto in cui non è adatto. Si tratta di una situazione molto simile a quella che riguarda la quarta e ultima categoria di *bias* (anche se, in verità, si tratta di qualcosa di diverso): gli algoritmi che, pur non presentando difetti nella progettazione e nell'utilizzo, finiscono per avere esiti discriminatori poiché riproducono *bias* esistenti nella società[495]. Così, un sistema addestrato su un *dataset* formato senza errori, avrà effetti nella pratica discriminatori se quei dati rappresentano vicende della vita reale che penalizzano determinate minoranze o categorie, e lo farà in modo ancor più draconiano qualora venisse impiegato nel contesto per cui è stato pensato, e in base a un algoritmo sviluppato perfettamente. Sarebbe il caso, ad esempio, di un sistema progettato per l'automazione delle procedure di selezione per posizioni lavorative dirigenziali in molti settori della nostra economia, che, se venisse sviluppato facendo riferimento alle scelte in tal senso prese da molte aziende nei decenni passati, finerebbe, in assenza di correttivi, per discriminare i candidati di sesso femminile, vista la scarsa presenza, nella società, di donne in posizioni apicali[496]. Tali discriminazioni, presenti nella società, verrebbero replicate attraverso l'algoritmo, con l'ulteriore rischio di renderne più difficile l'identificazione, a causa della già analizzata parvenza di oggettività che, in molti casi, circonda la tecnologia nella coscienza comune[497]. La particolare difficoltà nel percepire la discriminazione algoritmica, del resto, è un tratto comune a tutte le tipologie di *bias* analizzate fin qui. È anche per questo che tale nuovo tipo di discriminazione, dal punto di vista giuridico, pone questioni – anch'esse di indubbia novità - particolarmente pressanti per l'effettività del principio di eguaglianza, alle quali saranno dedicate le prossime pagine.

[495]Il tema è particolarmente dibattuto nella letteratura scientifica, compresa quella giuridica: cfr. ad es. J. KLEINBERG, J. LUDWIG, S. MULLAINATHAN, C. R. SUNSTEIN, *Discrimination in the age of algorithms*, in *Journal of legal analysis*, 10, 2018, p. 113-174; M. SELMI, *Algorithms, discrimination and the law*, *Ohio State Law Journal,* 82, 4, p. 611-652; E. NTOUTSI ET AL., *Bias in Data-driven Artificial Intelligence Systems—An Introductory Survey*, in*WIREs Data Mining and Knowledge Discovery*, 10, 3, 2020, https://onlinelibrary.wiley.com/doi/10.1002/widm.1356 (14 maggio 2022);A. E. R. PRINCE, D. SCHWARCZ,*Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, in *Iowa Law Review* 105, 3, 2020, p. 1257-1318 ss.;F. J. ZUIDERVEENBORGESIUS,*Strengthening legal protection against discrimination by algorithms and artificial intelligence*, in*The International Journal of Human Rights*, 24, 10,2020, p. 1572-1593; F. ZUIDERVEENBORGESIUS, *Discrimination, artificial intelligence, and algorithmic decision-making*, Council of Europe, Directorate General of Democracy, 2018, https://bit.ly/3N28YZ7 (14 maggio 2022).

[496]Cfr. A. KÖCHLING, M. C. WEHNER, *Discriminated by an algorithm: asystematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development*, in*Business Research*, 13, 3, p. 795–848; A. LAMBRECHT, C. TUCKER, *Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads*, in*Management Science*, 65, 7, 2019, p. 2966–2981.

[497] Sul punto, cfr. in particolare A. BONEZZI, M. OSTINELLI, *Can algorithms legitimize discrimination?*, in *Journal of Experimental Psychology: Applied, 27*, 2, p. 447–459; Y. BIGMAN, K. GRAY, A. WAYTZ, M. ARNESTAD, D. WILSON, *Algorithmic discrimination causes less moral outrage than human discrimination*, PsyArXiv; 2020,doi:10.31234/osf.io/m3nrp (14 maggio 2022); S. QUINTARELLI, F. COREA, F. FOSSA, A. LOREGGIA, S. SAPIENZA, *AI: profilietici cit.*, p. 197-198.

4. Il diverso valore del principio di eguaglianza nei confronti dei poteri pubblici e privati e le conseguenze in materia di discriminazione algoritmica

La struttura del principio di eguaglianza è fin troppo nota[498]. Le prime Costituzioni liberali si limitavano ad enunciare la c.d. *eguaglianza formale*, intesa, inizialmente, come semplice rigetto dei privilegi e delle differenze di status che avevano caratterizzato le epoche precedenti. A partire da questo nucleo essenziale si è fatta strada, poi, l'idea che situazioni differenti meritino trattamenti giuridici differenziati, e che solo in tal modo i cittadini possano dirsi effettivamente eguali di fronte alla legge[499]. La c.d. *eguaglianza sostanziale*, invece, è entrata nella tradizione costituzionale dell'Europa occidentale con lo sviluppo dello stato costituzionale di diritto, e in particolare della sua componente sociale, avvenuto nel secondo dopoguerra[500]. Il dibattito sull'effettivo contenuto dell'eguaglianza sostanziale è stato, nei decenni, piuttosto ricco, in primo luogo nella dottrina italiana[501]: è da tutti condiviso, in ogni caso, che il principio imponga ai poteri pubblici importanti obblighi positivi al fine di mitigare gli effetti delle diseguaglianze strutturalmente presenti nella società. Allo stesso tempo, si esclude che tali doveri – finalizzati alla c.d. *eguaglianza di partenza* – si spingano ad imporre allo stato il compito di redistribuire la ricchezza fino ad appianare ogni differenza patrimoniale (perseguendo l'*eguaglianza di risultato* propria della tradizione socialista)[502].

---

[498] Nel ribadire l'impossibilità di ripercorrere l'intero dibattito dottrinale, italiano e straniero, sul principio di uguaglianza, senz'animo di completezza si rinvia ai già citati C. ESPOSITO, *Eguaglianza e giustizia nell'art. 3 della Costituzione cit;* L. PALADIN, *Eguaglianza (diritto cost.) cit.*;*Il principio costituzionale d'eguaglianza cit.* A. S. AGRÒ, U. ROMAGNOLI, *Commento all'art. 3 cost. cit.*P. BISCARETTI DI RUFFIA, *Uguaglianza cit*;A. CERRI, *Uguaglianza cit.*; M. CARTABIA, T. VETTOR (A CURA DI), *Le ragioni dell'uguaglianza cit*; M. LUCIANI,*I principi di eguaglianza e non discriminazione cit.*;T. F. BOTTS, *For equals only: race, equality and the equal protection clause cit.*; F. M. SOUCRAMANIEN, *Le principe d''égalité cit.*; J. H. WILKINSON, *The Supreme Court, the Equal Protection Clause cit.*;R. B. GINZBURG, *Interpretations of the equalprotectionclause cit.*;C. F. J. DOEBBLER, *Principle of non-discrimination cit.*

[499] Cfr. ad es. F. SORRENTINO, *Eguaglianza formale*, in *Costituzionalismo.it*, 3, 2017; D. FLORENZANO, *Il principio costituzionale di eguaglianza cit.*, p. 133 ss.

[500] Tra gli studi specifici in materia di eguaglianza sostanziale possono indicarsi, senz'animo di completezza, S. CASSESE, *L'eguaglianza sostanziale nella Costituzione: genesi di una norma rivoluzionaria*, in *Le carte e la storia*, 1, 2017, p. 5-13; B. CARAVITA, *Oltre l'uguaglianza formale. Un'analisi dell'art. 3 c. 2 della Costituzione*, Padova, 1984;A. GIORGIS, *Art. 3 c. 2*, in R. BIFULCO, A, CELOTTO, *Commentario alla Costituzione*, I, Torino, 2006, p. 88-113; A. GIORGIS, *La costituzionalizzazione dei diritti all'eguaglianza sostanziale*, Napoli, 1999; M. C. GIORGI, *L'uguaglianza sostanziale nel lungo dibattito costituente*, in *Rivista Trimestrale di Diritto Pubblico*, 1, 2018, p. 9-43; I. SANNA (a cura di), *Diritto di cittadinanza e uguaglianza sostanziale*, Roma, 2014; A. D'ALOIA, *Eguaglianza sostanziale e diritto diseguale. Contributo allo studio delle azioni positive nella prospettiva costituzionale*, Padova, 2002; M. F. DE TULLIO, *Uguaglianza sostanziale e nuove dimensioni della partecipazione politica*, Napoli, 2020.

[501] La divisione, nei primi anni di vigenza della Carta, è stata in primo luogo tra chi tendeva a sminuire l'effettività del principio, attribuendogli valore prevalentemente programmatico, e chi (si veda, ad esempio, l'opera di Livio Paladin) ne sottolineava l'indubbio contenuto precettivo, in grado di ispirare l'intero ordinamento (interpretazione oggi pacifica). Per una ricostruzione cfr. D. FLORENZANO, *Il principio costituzionale di eguaglianza cit.*, p. 155 ss.

[502] L'idea che dal principio di eguaglianza come codificato all'art. 3 della Carta sia possibile ricavare che l'eguaglianza di risultato rappresenti una delle finalità dell'ordinamento – probabilmente presente, per la verità, almeno nelle convinzioni dei costituenti di matrice socialista e comunista - si scontra irrimediabilmente con altre dispozioni costituzionali, in primo luogo con quelle (artt. 41 e 42 Cost.) che tutelano l'iniziativa economica e la proprietà privata. Come osservato da più parti in dottrina l'eguaglianza di partenza ha rappresentato la felice sintesi delle diverse anime che formavano l'Assemblea costituente, consentendo a ciascuna di esse di identificarsi nella disposizione senza snaturare l'apparato ideologico di riferimento. Cfr. in particolare A. PIZZORUSSO, *Che cos'è l'eguaglianza, il principio*

La costituzionalizzazione dello stato sociale, com'è noto, è una delle principali differenziazioni tra le tradizioni costituzionali europee e quella degli Stati Uniti d'America, rimasta più aderente all'originario modello dello stato liberale di diritto. Il principio di eguaglianza è entrato nella Costituzione americana con l'introduzione, al termine della guerra civile, del XIV Emendamento, abolitivo della schiavitù[503]. L'imposizione ai poteri pubblici di doveri positivi per il raggiungimento dell'eguaglianza nelle possibilità tra i cittadini è stata ed è oggetto, nell'ordinamento americano, di un vivace dibattito dottrinale e giurisprudenziale[504]. È un dato di realtà, inoltre, che, nonostante profonde differenze in materia tra gli stati della federazione, negli Stati Uniti non esista un *welfare state* paragonabile al modello europeo, e, anzi, la spesa pubblica necessaria a metterlo in atto sarebbe da più parti accusata di incostituzionalità[505]. Ciò nonostante, la Corte Suprema ha più volte affermato la legittimità costituzionale delle c.d. *affermative action*, politiche ispirate al principio di eguaglianza sostanziale, introduttive di un trattamento differenziato per determinati gruppi sociali, al fine di colmare differenze strutturali presenti nella società. Ciò è avvenuto, in particolare, per provvedimenti federali volti a mitigare le disparità causate dalla segregazione razziale[506].

Un tratto comune del principio di eguaglianza negli ordinamenti democratici, invece, è la sua valenza generale unicamente verticale, nei rapporti tra individuo e autorità. L'eguaglianza funge da vincolo per il Legislatore - che non può, come abbiamo visto, assumere atteggiamenti discriminatori

---

etico e la norma giuridica nella vita reale, Roma, 1983, che descrive nell'art. 3 c. 2 il punto d'arrivo della ricerca di un termine medio tra l'egalitarismo ("a ciascuno secondo i suoi bisogni") e l'eguaglianza formale ("a ciascuno secondo i propri meriti"). V. anche S. CASSESE, *L'eguaglianza sostanziale nella Costituzione cit.*; D. FLORENZANO, *Il principio costituzionale di eguaglianza cit.*, p. 155 ss.

[503] Sul XIV Emendamento, e più in generale sulla concezione statunitense dell'eguaglianza, cfr. *ex multis* S. ENGDAHL (a cura di), *Amendment XIV: Equal Protection*, in *Constitutional amendments: beyond the bill of rights*,Farmington Hills, 2009; K. T. LASH, *The Fourteenth Amendment and the privileges and immunities of American citizenship*, Cambridge, 2014; M. K. CURTIS, *No State shall abridge. The XIV Amendment and the Bill of Rights*, Durham (USA), 1986; D. L. HUDSON, *The Fourteenth Amendment. Equal protection under the law*, Berkeley, 2002; S. VERBA, G. R. ORREN, *The meaning of equality in America*, in *Political Science Quarterly*, 100, 3, 1985, p. 369-387; J. P. ROCHE, *Equality in America: the expansion of a concept*, in *North Carolina Law Review*, 43, 2, 2, 1969, p. 249-270.

[504] J. E. KELLUOGH, *Understanding affirmative action*, Washington, 2006; P. J. MISHKIN, *The uses of ambivalence: reflections of the Supreme Court and the constitutionality of affirmative action*, in *University of Pennsylvania Law Review*, 131, 1983, p. 107 ss.; R. J. FISCUS, *The constitutional logic of affirmative action*, Durham-Londra, 1992; M. ROSENFELD, *Affirmative action, justice, and equalities: a philosophical and constitutional appraisal*, in *Ohio State Law Review*, 46, 1985, p. 845 ss.; E. S. ANDERSON, *Integration, affirmative action and strict scrutiny*, in *New York University Law Review*, 77, 5, 2002, p. 1195-1271.

[505] Il tema dello spazio concesso ai poteri pubblici in materia di politiche distributive e spesa sociale negli Stati Uniti è complesso e risalente almeno agli anni del *New Deal*, le cui politiche furono oggetto di diverse pronunce, di segno alterno, da parte della Corte Suprema. Il dibattito prosegue ancora oggi, ad esempio riguardo al *Patient Protection and Affordable Care Act*, la riforma dell'assistenza sanitaria introdotta nel 2012 dall'amministrazione Obama, più volte portata all'attenzione della Corte, che si è, finora, sempre espressa per la sua compatibilità con la Costituzione, cfr. *National Federation of Independent Business v. Sebelius*, 567 U.S. 519 (2012). Non può trascurarsi, inoltre, l'incidenza delle profonde differenze, in materia di concezione delle politiche distributive, tra i distinti ordinamenti federati degli Stati Uniti.In letteratura v. ad es. S. KRISLOV, *American Welfare Policy and the Supreme Court*, in *Current History*, 65, 383, 1973, p. 33–42.

[506] Cfr., ad esempio, *Regents of the University of California v. Bakke*, 438 U.S. 265 (1978); *Grutter v. Bollinger*, 539 U.S. 306 (2003); *Fisher v. University of Texas*, 570 U.S. 297 (2013) in cui la Corte Suprema ha ritenuto costituzionale l'utilizzo dell'etnia tra i fattori di decisione nella selezione degli studenti ammessi a un corso universitario, pur col vaglio rigoroso dello *strict scrutiny* su tali politiche.

– e per l'Autorità amministrativa, la quale, nell'esercizio della propria discrezionalità, è comunque tenuta ad evitare differenziazioni immotivate (la circostanza, del resto, non è che la consequenza della sottoposizione a una legge che a tale principio si ispira). I privati, normalmente, non sono obbligati a tenere in considerazione, nelle loro azioni, le dinamiche dell'eguaglianza, e possono ispirare il proprio comportamento a criteri che sarebbero considerati discriminatori se applicati dai poteri pubblici. Un individuo è libero di scegliere di accompagnarsi solamente con persone che presentino certe caratteristiche fisiche, o appartengano a una determinata religione; chi esercita un'attività commerciale ha ampia libertà per quando riguarda i criteri di scelta dei propri clienti; un datore di lavoro non è tenuto al rispetto di un criterio generalizzato di eguaglianza di partenza per quanto riguarda le assunzioni. In breve, il principio di eguaglianza non rappresenta un limite generale all'autonomia privata.[507]

Quanto appena affermato, però, è vero solo con importanti distinguo. Esigenze di tutela dell'eguaglianza, infatti, hanno portato allo sviluppo di importanti vincoli all'autonomia privata. Alcune possibili ragioni di discriminazione appaiono particolarmente odiose e i legislatori di vari paesi hanno scelto di impedire ai privati di metterle in atto. Pur con importanti differenze, ciò è avvenuto con profonde similitudini tra ordinamenti distinti e, in particolare, tra i paesi membri dell'Unione Europea – il cui operato ha avuto, come subito si dirà, un ruolo decisivo – e gli Stati Uniti. Sullo scenario europeo, fin dagli anni '70 una lunga serie di direttive è intervenuta al fine di uniformare le normative frammentarie degli stati membri, dando origine al corpo di norme che oggi

[507] Dell'inidoneità del principio di eguaglianza a rappresentare un criterio generale di valutazione dell'autonomia privata si è estensivamente occupata, in Italia, sia la dottrina pubblicistica che quella privatistica. Non si può trascurare, in ogni caso, il ruolo sempre più penetrante che l'eguaglianza gioca nei rapporti tra privati, in dialogo col principio di autodeterminazione e con la protezione della dignità, nei settori afferenti al c.d. diritto antidiscriminatorio di matrice eurounitaria (tema che sarà analizzato nelle pagine seguenti) e nei sempre più numerosi presidi normativi a tutela del contraente debole nel diritto dei contratti, come evidenziato da parte della dottrina civilistica più recente. In particolare cfr., da vari punti di vista, F. VARI, *L'affermazione del principio di eguaglianza nei rapporti tra privati. Profili costituzionali*, Torino, 2017; E. NAVARETTA, *Principio di uguaglianza, principio di non discriminazione e contratto*, in *Rivista di diritto civile*, 2014, p. 547-566; G. CARAPELLA FIGLIA, *Il divieto di discriminazione quale limite all'autonomia contrattuale*, in *Rivista di diritto civile*, 2015, p. 1387-1418; A. GENTILI, *Il principio di non discriminazione nei rapporti civili*, in *Rivista critica del diritto privato*, 2, 2009, p. 207-231; P. RESCIGNO, *Il principio di eguaglianza nel diritto privato*, in *Rivista trimestrale dir. proc. civ.*, 1959, p. 1515 ss.; P. RESCIGNO, *Sul cosiddetto principio d'uguaglianza nel diritto privato*, in *Foro it.*, 1959, p. 664 ss.; D. CARUSI,*Principio di eguaglianza, diritto singolare e privilegio.Rileggendo i saggi di Pietro Rescigno*, Napoli, 1998. Del tema si è occupata anche la dottrina tedesca, nella quale, peraltro, l'introduzione di normativa antidiscriminatoria è stata, in passato, accompagnato da critiche feroci, cfr. N. VENNEMAN, *The German Draft Legislation On the Prevention of Discrimination in the Private Sector*, in *German Law Journal*, 3, 3, 2002; K. H. LAUDER, *The German Proposal of an "Anti-Discrimination"-Law: Anticostitutional and Anti-Common Sense. A Response to Nicola Venneman*, in *German Law Journal*, 3, 5, 2002. Il tema di un'eventuale efficacia orizzontale del principio di eguaglianza, invece, sostanzialmente non esiste negli Stati Uniti, in cui la valenza unicamente verticale della Costituzione, e in particolare del *Bill of Rights*, sancita dalla c.d. *state actiondoctrine*non è mai stata messa in discussione, ed è lo stesso testo del XIV Emendamento a identificare esplicitamente i poteri pubblici quali unici destinatari della norma («*no states shall...*»), cfr. S. JAGGI, *State action doctrine*, in *Max Planck Encyclopedia of Comparative Constitutional Law*, 2017, https://oxcon.ouplaw.com/view/10.1093/law-mpeccol/law-mpeccol-e473 (20 luglio 2021); T. PERETTI, *Constructing the State Action Doctrine, 1940–1990*, in *Law & Social Inquiry*, 35, 2, 2010, p. 273 ss. Ciò non ha escluso, come vedremo, l'emanazione anche oltreoceano di discipline puntuali che limitano l'autonomia privata a protezione dell'individuo da fattori di discriminazione particolarmente odiosi.

compongono il c.d. *diritto antidiscriminatorio*[508]. Le iniziative delle istituzioni comunitarie si sono concentrate, nella fase iniziale, sulla creazione di un mercato del lavoro che promuovesse un'effettiva occupazione femminile, proibendo e sanzionando ogni genere di discriminazione in base al sesso sul lavoro, dal momento dell'assunzione a quello del pensionamento[509]. In una seconda fase, a partire dal principio degli anni 2000, le direttive dell'Unione in materia hanno preso in considerazione altri fattori di discriminazione particolarmente odiosi – e in particolare la "razza" e l'origine etnica, l'età, l'orientamento sessuale, le convizioni personali, la religione, la disabilità – e ambiti di applicazione diversi dal contesto lavorativo, come l'accesso a determinati servizi essenziali, quali sanità, protezione e assistenza sociale, istruzione[510]. Un'evoluzione simile è avvenuta al di là dell'Atlantico, con l'emanazione, nel 1964, del noto *Civil Rights Act*, che pose fine ad ogni residuo del regime di segregazione, proibendo ogni discriminazione in base a etnia, religione, sesso o origini nazionali nella registrazione ai fini del voto, nell'istruzione, sul lavoro e nell'accesso ai luoghi aperti al pubblico[511]. Il testo è stato, nei decenni successivi, più volte emendato, con l'aggiunta di nuovi fattori di discriminazione protetti e ulteriori campi d'applicazione. Nel 1968, in particolare, furono introdotte speciali prerogative a tutela delle tribù di nativi americani (col c.d. *Indian Civil Rights Act*)[512] e la proibizione di ogni discriminazione nell'accesso alla locazione o alla proprietà immobiliare (c.d. *Fair Housing Act*)[513]. Nel 1990,

[508]L. CALAFÀ, D. GOTTARDI, *Il diritto antidiscriminatorio tra teoria e prassi applicativa*, Roma, 2009; M. BARBERA, M. AIMO, *Il nuovo diritto antidiscriminatorio: il quadro comunitario e nazionale*, Milano, 2007; E. EVELYN, P. WATSON, *EU Anti-Discrimination Law*, Oxford, 2012;E. CONSIGLIO, *Che cos'è la discriminazione? Un'introduzione teorica al diritto antidiscriminatorio*, Torino, 2020;M. BARBERA, A. GUARISO (a cura di), *La tutela antidiscriminatoria. Fonti strumenti interpreti*, Torino, 2020.

[509] Per questa prima fase cfr., in particolare, Dir. 75/117/CEE, *per il ravvicinamento delle legislazioni degli Stati Membri relative all'applicazione del principio della parità delle retribuzioni tra i lavoratori di sesso maschile e quelli di sesso femminile*; Dir. 76/207/CEE, *relativa all'attuazione del principio della parità di trattamento fra gli uomini e le donne per quanto riguarda l'accesso al lavoro, alla formazione e alla promozione professionali e le condizioni di lavoro*; Dir. 86/378/CEE, *relativa all'attuazione del principio della parità di trattamento tra gli uomini e le donne nel settore dei regimi professionali di sicurezza sociale Il Consiglio delle Comunità europee*; Dir. 97/80/CE, *riguardante l'onere della prova nei casi di discriminazione basata sul sesso*; Dir. 86/613/CEE, *relativa all'applicazione del principio della parità di trattamento fra gli uomini e le donne che esercitano un'attività autonoma, ivi comprese le attività nel settore agricolo, e relativa altresì alla tutela della maternità*; Dir. 92/85/CEE, *concernente l'attuazione di misure volte a promuovere il miglioramento della sicurezza e della salute sul lavoro delle lavoratrici gestanti, puerpere o in periodo di allattamento*. Per una ricostruzione teorica cfr. E. EVELYN, P. WATSON, *EU Anti-Discrimination Law cit.*, p. 180 ss.

[510] Cfr., per questa seconda fase, in particolare, la Dir. 2000/43/CE (c.d. direttiva razza) *che attua il principio della parità di trattamento fra le persone indipendentemente dalla razza e dall'origine etnica* e la Dir. 2000/78/CE (c.d. direttiva quadro) *che stabilisce un quadro generale per la parità di trattamento in materia di occupazione e di condizioni di lavoro*. In dottrina cfr. in particolare M. BARBERA, M. AIMO, *Il nuovo diritto antidiscriminatorio cit.*, p. XIX ss.

[511] U.S. *Civil Rights Act* of 1964, Pub. L. 88-532, 78 Stat. 241, 2 luglio 1964. Per dei commenti in letteratura cfr. *ex multis* R. D. LOEVY, *The Civil Rights Act of 1964: The Passage of the Law That Ended Racial Segregation*, Albany, 1997; H. GRAHAM, *The Civil Rights Era: Origins and Development of National Policy 1960–1972*, New York, 1990; R. F. GREGORY, *The Civil Rights Act and the Battle to End Workplace Discrimination*, Lanham, 2014; D. B. RODRIGUEZ, B. R. WEINGAST, *The Positive Political Theory of Legislative History: New Perspectives on the 1964 Civil Rights Act and Its Interpretation*, in *University of Pennsylvania Law Review*, 151, 2003, p. 1417-1542.

[512] Si tratta dei Titoli da II a VII del U.S. *Civil Rights Act* of 1968, Pub. L. 90-284, 82 Stat. 73, 11 aprile 1968.

[513] Titoli VIII e IX del menzionato *Civil Rights Act* del 1968.

invece, l'*Americans with Disabilities Act* incluse la disabilità tra i fattori di discriminazione vietati[514]. Parallelamente, un contributo decisivo allo sviluppo della normativa antidiscriminatoria statunitense venne dalla giurisprudenza, che nel corso degli anni ne ha allargato il campo di applicazione (estendendolo, ad esempio, alle discriminazioni sulla base dell'orientamento sessuale)[515].

Completano, infine, il quadro del diritto antidiscriminatorio un buon numero di convenzioni internazionali, diverse delle quali elaborate in seno all'OIL, e vari presidi penalistici che mutano di molto tra i vari ordinamenti nazionali (si pensi, per quanto riguarda il caso italiano, alla c.d. Legge Mancino)[516].

In breve, il principio di eguaglianza proibisce ai poteri pubblici ogni trattamento diseguale che non appaia giustificato secondo un canone di ragionevolezza. Gli operatori privati, invece, a prescindere dalle loro dimensioni e dalla loro concreta capacità di influenza, sono tenuti unicamente ad evitare talune condotte discriminatorie specificamente individuate, per la loro gravità, dal legislatore statale o sovranazionale. Ciò ha importanti conseguenze sull'analisi della discriminazione algoritmica. Non può ignorarsi, infatti, che in molti casi l'intelligenza artificiale si innesta in meccanismi decisionali di privati, spesso in situazioni particolarmente delicate per l'individuo (com'è il caso, più volte richiamato, della selezione per un posto di lavoro) né che è privata, nella stragrande dei casi, la proprietà intellettuale di tali tecnologie. La riflessione sulla discriminazione algoritmica deve, dunque, essere inquadrata all'interno dell'ampia differenza di disciplina che caratterizza, in materia di eguaglianza, poteri pubblici e privati. Preme evidenziare, in particolare, due punti che saranno esaminati più in profondità nei prossimi paragrafi, al momento dell'analisi delle possibili prospettive di regolazione del fenomeno.

In primo luogo, come già detto, la presenza di algoritmi rende più difficoltosa la stessa identificazione delle discriminazioni, a causa della già analizzata percezione di oggettività che spesso accompagna la tecnologia, del possibile intervento della c.d. distorsione dell'automazione, della difficoltà ad identificare eventuali bias e, talvolta, a comprendere lo stesso funzionamento generale dell'algoritmo, per ragioni tecniche o per la presenza di un segreto industriale. Ciò può

[514] U.S. *American with Disabilities Act* of 1990, 42 U.S.C. par. 12101, 26 luglio 1990.

[515] Cfr. in particolare le recenti sentenze della Corte Suprema nei casi*Bostock v. Clayton County* (2020) e *Altitude Express, Inc. v. Zarda* (2020), che hanno riconosciuto la protezione prevista dal *CivilRightsAct* per le discriminazioni in base al sesso a quelle per l'orientamento sessuale, e *R.G. & G.R. Harris Funeral Homes Inc. v. Equal Employment Opportunity Commission* (2020), che lo ha fatto per le discriminazioni in ragione dell'identità di genere.

[516] Cfr. ad es. la Convenzione OIL n. 100, *sull'uguaglianza di retribuzione fra mano d'opera maschile e mano d'opera femminile per un lavoro di valore uguale*, 1951 e la Convenzione OIL n. 111, *sulla discriminazione in materia di impiego e nelle professioni*, 1958. La c.d. "legge Mancino", dal nome dell'onorevole primo firmatario, è la Legge n. 205 del 25 giugno 1993, *Conversione in legge, con modificazioni, del decreto-legge 26 aprile 1993, n. 122, recante misure urgenti in materia di discriminazione razziale, etnica e religiosa*, in G.U. n. 148 del 26 giugno 1993. Per un quadro completo delle fonti del diritto antidiscriminatorio si rimanda a F. BILOTTA, A. ZILLI, *Codice di diritto antidiscriminatorio*. Ospedaletto (PI), 2019.

rendere particolarmente difficile risalire alla reale ragione di un trattamento differenziato, un'esigenza fondamentale, come abbiamo visto, per l'applicazione delle norme di diritto antidiscriminatorio. L'operatore privato che utilizzasse, nei propri meccanismi decisionali, tecnologie che conducono a tali esiti discriminatori potrebbe tentare di far apparire tali risultati come giustificati da ragioni oggettive, grazie allo "schermo" della tecnologia, più di quanto non avvenga oggi. L'ipotesi, non del tutto improbabile, di una sua eventuale inconsapevolezza dei bias del sistema, e dunque perfetta buonafede, aggiunge ulteriori elementi di complessità[517].

In secondo luogo, il ruolo crescente dell'intelligenza artificiale nei processi decisionali, e in particolare delle sue versioni in cui l'analisi dei dati ha un ruolo prevalente, rischia di far perdere di senso l'identificazione dei fattori di discriminazione particolarmente odiosi che i legislatori degli ordinamenti democratici precludono, come abbiamo visto, anche ai privati. Non vi sono particolari ostacoli per la formazione di dataset per l'addestramento in cui i dati relativi a etnia, orientamento sessuale, età, ecc. non sono presenti, né, in generale, allo sviluppo di algoritmi che non tengano in considerazione queste variabili. Tali fattori di discriminazione, però, rischiano di ripresentarsi sotto diverse spoglie. Infatti, diverse altre informazioni (c.d. *proxies*), apparentemente neutre, possono avere un impatto sproporzionato su determinate categorie o minoranze[518]. Ad esempio, l'area di residenza, la frequenza di alcune scuole, determinate abitudini di consumo possono essere indicatori significativi della probabile appartenza a una certa etnia, fascia d'età o livello di reddito. L'elaborazione su larga scala di dati di questo tipo rende tale probabilità prossima alla certezza. Un algoritmo basato sull'apprendimento automatico, dunque, se addestrato su un *set* di decisioni viziate da pregiudizi e discriminazioni, rischierà di replicare tali caratteristiche, anche qualora nella selezione dei dati di addestramento si facesse particolare attenzione ad escludere ogni identificativo prevedibile dell'appartenenza a gruppi svantaggiati. Quest'ultima condizione, infatti, rimarrebbe, nella larga maggioranza dei casi, rilevabile dalle informazioni rimanenti, in un modo difficilmente prevedibile e controllabile (posto che è proprio la particolare profondità e vastità dell'elaborazione dei dati a rendere desiderabile l'uso di algoritmi). Il pregiudizio presente nelle decisioni umane prese a modello, così, finirebbe comunque per essere replicato dall'algoritmo. Tali discriminazioni potrebbero, anzi, venire acuite dall'elaborazione dei dati con strumenti di intelligenza artificiale: è la critica mossa a diversi strumenti di polizia predittiva, accusati di concentrare l'attività della polizia nelle aree considerate a rischio maggiore, portando questa ad individuare un numero di reati

[517] Sul punto, cfr. in particolareK. MARTIN, *Algorithmic Bias and Corporate Responsibility: How companies hide behind the false veil of the technological imperative*, in K. MARTIN (a cura di), *Ethics of data and analytics*, Boca Raton, 2022.

[518]A. E. PRINCE, D. SCHWARCZ, *Proxy discrimination in the age of artificial intelligence and big data*, in *Iowa Law Review*, 105, 2019, p. 1257 ss.; R. BENJAMIN, *Assessing Risk, Automating Racism*, in *Science*, 366, 6464, p. 421–422; J. KLEINBERG, J. LUDWIG, S. MULLAINATHAN, C. R. SUNSTEIN, *Discrimination in the age of algorithms*, in *Journal of legal analysis*, 10, 2018, p. 137 ss.

sempre più alto e a trascurare la criminalità operante in altri luoghi, generando dati, magari utilizzati per l'autoapprendimento, ancora più negativi per quelle zone (e, di conseguenza, connotati di un particolare stigma per le persone che li abitano o frequentano, magari accomunate dall'appartenenza a minoranze etniche marginalizzate)[519]. La strategia normativa perseguita finora, che pone, come limite generale all'autonomia privata, il solo divieto di discriminazioni che si basino su determinati fattori, considerati particolarmente odiosi per ragioni storiche e sociali, rischia, dunque, di rilevarsi impotente di fronte all'utilizzo, da parte di poteri privati sempre più influenti, di strumenti di intelligenza artificiale nei meccanismi decisionali, che potrebbero portare, attraverso l'analisi di dati all'apparenza neutri, a effetti discriminatori equivalenti a quelli basati sui fattori vietati.

5. I primi casi di discriminazione algoritimica affrontati dalle corti

Per ora, non sembrano molti i casi di discriminazione algoritmica giunti in tribunale. Quelli particolarmente noti e commentati in dottrina, inoltre, si contano sulle dita di una mano. Tra questi, i più celebri sono, forse, due casi provenienti dal continente nordamericano, *Loomis v. Wisconsin*[520] ed *Ewert v. Canada*[521]. Entrambe le vicende scaturivano dall'utilizzo di tecnologie avanzate nella valutazione della pericolosità sociale e della possibilità di recidiva in seno a procedimenti penali, e

[519]Sulle tecnologie di intelligenza artificiale applicate all'attività di *law enforcement*, e le loro possibili conseguenze sui diritti fondamentali cfr., da prospettive differenti, A. MEIJER, M. WESSELS, *Predictive policing: review of benefits and drawbacks*, in *International Journal of Public Administration*, 42, 12, 2019, p. 1031 ss.; W. PERRY, B. MCINNIS, C. C. PRICE, S. C. SMITH, J. S. HOLLYWOOD, *Predictive policing. The role of crime forecasting in law enforcement operations*, Washington, 2013. V. ad es. S. QUATTROCOLO, *Equo processo penale e sfide della società algoritmica*, in *Rivista di BioDiritto – BioLaw Journal*, 1, 2019, p. 135 ss. e S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs. rischi e paure della giustizia digitale "predittiva"*, in *Cassazione penale*, 59, 4, 2019, p. 1748 ss.; B. PEREGO, *Predictive policing: trasparenza degli algoritmi, impatto sulla privacy e risvolti discriminatori*, in *BioLaw Journal - Rivista di BioDiritto*, 2, 2020, p. 447-465; F. BASILE, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, in F. BASILE, M. CATERINI, S. ROMANO (a cura di), *Il sistema penale ai confini delle hard sciences. Percorsi epistemologici tra neuroscienze e intelligenza artificiale*, Ospedaletto (PI), 2021; K. ALIKHADEMI, E. DROBINA, D. PRIOLEAU, *A review of predictive policing from the perspective of fairness*, in *Artificial Intelligence and Law*,30, p. 1–17, 2022, https://doi.org/10.1007/s10506-021-09286-4; J. L. M. MCDANIEL, K. G. PEASE, *Predictive policing and artificial intelligence*, 2021, New York.

[520]*State v. Loomis*, 881 N.W.2d 749 (Wis. 2016), con nota in *Harvard Law Review*, 130, 2017, p. 1530 ss. Per alcunicommenti, v. L. HAN-WEI, L. CHING-FU, C. YU-JIE, *Beyond State v Loomis: Artificial Intelligence, Government Algorithmization and Accountability*, in *International Journal of Law and Information Technology*, 27, 2, p. 122–141; I. DE MIGUEL BERIAIN, *Does the Use of Risk Assessments in Sentences Respect the Right to Due Process? A Critical Analysis of the Wisconsin v. Loomis Ruling*, in *Law, Probability and Risk*, 17, 1, p. 45–53; A.L. WASHINGTON, *How to Argue with an Algorithm: Lessons from the COMPAS-ProPublica Debate*, in *Colorado Technology Law Journal*, 17, 2018, p. 131 ss.; J. LIGHTBOURNE, *Damned Lies & Criminal Sentencing Using Evidence-Based Tools*, in *Duke Law & Technology Review*, 15, 2017, p. 327 ss.; K. FREEMAN, *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis*, in *North Carolina Journal of Law & Technology*, 18, 5, 2016, p. 75 ss.

[521]*Ewert v. Canada*, 2018 SCC 30 [2018] 2 S.C.R. 165. In letteraturacfr. A.M. HAAG, A. BOYES, J. CHENG, A. MACNEIL, R. WIROVE, *An introduction to the issues of cross-cultural assessment inspired by Ewert v. Canada*, in *Journal of Threat Assessment and Management,* 3, 2, 2016, p. 65–75; D. G. KRONER, *The Ewert v. Canada judgment: Moving forward*, in *Journal of Threat Assessment and Management,* 3, 2, 2016, p. 122–127; M. E. OLVER, *Some considerations on the use of actuarial and related forensic measures with diverse correctional populations*, in *Journal of Threat Assessment and Management,* 3, 2, 2016, p. 107–121; E. HILL, J. WOLFE, *Ewert v. Canada: Shining Light on Corrections and Indigenous People*, in *The Supreme Court Law Review: OsgoodÈs Annual Constitutional Cases Conference* , 94, 15, 2020, p. 391-413.

hanno acquisito particolare notorietà per l'importanza dei diritti interessati – le decisioni riguardavano la libertà personale dei soggetti coinvolti – e perché l'algoritmo sembrava replicare fattori di discriminazione particolarmente odiosi, come l'appartenza a determinate minoranze etniche storicamente svantaggiate. Come si dirà, i due casi hanno avuto esiti opposti. Allo stesso tempo, pur in misura molto diversa, entrambe le corti – come non era mai avvenuto prima da parte di un organo giurisdizionale - hanno mostrato consapevolezza dell'attualità e dell'urgenza dei problemi posti dall'intreccio tra discriminazioni storiche e nuove tecnologie.

L'imputato nel primo dei processi citati, Eric Loomis, era un piccolo criminale accusato di aver preso parte a uno scontro a fuoco avvenuto nel 2013 nella città di La Crosse. In seguito alla sua ammissione di colpevolezza per alcuni dei capi d'imputazione, il *Wisconsin Department of Corrections* aveva predisposto un c.d. *presentencing investigation report* (PSI) che conteneva, oltre a una relazione sui precedenti, i risultati dell'elaborazione del suo profilo col software COMPAS (acronimo di *Correctional Offender Management Profiling for Alternative Sanctions*). La giuria, anche in base a tale valutazione della pericolosità dell'imputato, aveva condannato Loomis a sei anni di reclusione e ulteriori 5 di sorveglianza speciale. Completava il quadro la circostanza che la metodologia e il funzionamento concreto del sofware COMPAS fossero coperti da segreto industriale.

Loomis presentò una mozione contro l'esecuzione della condanna, sostenendo che l'uso dell'algoritmo nella commisurazione della pena violava il proprio diritto a una sentenza frutto di una valutazione individualizzata, adeguatamente motivata e fondata su informazioni corrette. Nell'istanza, inoltre, sosteneva che la corte avesse illegittimamente preso in considerazione, nell'infliggere una pena così severa, il suo genere (presente tra i dati elaborati da COMPAS)[522]. Sia la corte di primo grado che la Corte Suprema del Wisconsin, successivamente adita da Loomis, rifiutarono le sue argomentazioni e confermarono la condanna. I giudici, in particolare, sostennero che stesse a Loomis dimostrare che il genere era stato un elemento decisivo nella decisione che aveva portato alla sua condanna e che egli avrebbe potuto contestare, ed eventualmente correggere, eventuali informazioni scorrette contenute nel PSI[523]. Riguardo all'individualizzazione della sentenza, la Corte Suprema del Wisconsin riconobbe, in accordo con Loomis, che le valutazioni di COMPAS non potevano mai riferirsi al singolo, ma semplicemente a un gruppo indeterminato di individui che condividevano determinate caratteristiche. Ciò nonostante, secondo la massima autorità giudiziaria del Wisconsin la personalizzazione del giudizio sull'imputato era assicurata dalla possibilità, per i giudici, di valutare criticamente i risultati del sistema, che si inserivano in una

[522] Cfr. *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016), p. 756-757.
[523] *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016), p. 760-761 e 765-767.

valutazione del rischio più ampia, riassunta in un report apposito, dal quale la corte poteva, in ogni caso, sempre discostarsi [524] .Successivamente, la Corte Suprema federale, di fronte alla quale Loomisappellò la decisione presa dai giudici del Wisconsin, confermò le decisioni precedenti, respingendo l'appello senza istruire la causa[525].

La vicenda è stata ulteriormente complicata da una ricerca sul funzionamento di COMPAS condotta dall'ONG americana ProPublica negli stessi anni del caso *Loomis* (2013-2016) che ha avuto l'effetto di far crescere enormemente l'interesse per questo genere di software e le polemiche attorno al loro utilizzo[526]. Analizzando i risultati di COMPAS su larga scala, infatti, ProPublica riscontrò un'eccessiva severità nelle sue valutazioni – ad esempio, appena il 20% di chi era stato designato come possibile colpevole di delitti violenti li aveva, poi, effettivamente commessi – e un *bias*che sfavoriva gli imputati afroamericani e ispanici, assegnando loro, a parità di storia criminale, un rischio di recidiva più alto. Nonostante l'algoritmo sia coperto, come già detto, da segreto industriale, è facile ipotizzare che ciò non sia che la replica del pregiudizio razziale che ha caratterizzato, per lungo tempo, il sistema giudiziario americano, i cui precedenti sono stati probabilmente utilizzati per sviluppare l'algoritmo, senza l'introduzione di sistemi di mitigazione di tale *bias* efficaci.

La vicenda alla base del caso *Ewert v. Canada* è, per molti versi, simile. Mr. Ewert è rinchiuso da oltre 30 anni nelle prigioni federali canadesi, in cui sta scontando due ergastoli, rispettivamente per omicidio consumato e tentato, ed appartiene, etnicamente e culturalmente, a uno dei popoli indigeni del Paese. Nel 2017 ha contestato l'utilizzo da parte del *Correctional Service of Canada* (CSC) di cinque diversi strumenti per la profilazione psicologica e la valutazione del rischio di pericolosità sociale nelle decisioni riguardanti il suo regime di detenzione e la possibilità di accedere a forme di esecuzione penale extramuraria. È di particolare interesse, ai nostri fini, il principale argomento utilizzato da Ewert: le tecnologie in questione, essendo state "addestrate" e utilizzate prevalentemente per valutazioni relative a persone non appartenenti a minoranze indigene, non

[524]«If a COMPAS risk assessment were the determinative factor considered at sentencing this would raise due process challenges regarding whether a defendant received an individualized sentence. As the defense expert testified at the post-conviction motion hearing, COMPAS is designed to assess group data. He explained that COMPAS can be analogized to insurance actuarial risk assessments, which identify risk among groups of drivers and allocate resources accordingly; Just as corrections staff should disregard risk scores that are inconsistent with other factors, we expect that circuit courts will exercise discretion when assessing a COMPAS risk score with respect to each individual defendant. [Ultimately, we disagree with Loomis because consideration of a COMPAS risk assessment at sentencing along with other supporting factors is helpful in providing the sentencing court with as much information as possible in order to arrive at an individualized sentence]», *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016), p. 764-765.

[525]*Loomis v. Wisconsin*,137 S.Ct. 2290 (2017).

[526]J. ANGWIN, J. LARSON, S. MATTU, L. KIRCHNER, *Machine Bias. There're software used across the country to predict future criminals. And it's biased against blacks*, ProPublica, 23 maggio 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (20maggio 2022). La risposta di Northpointe, l'azienda titolare della proprietà intellettuale di COMPAS, è disponibile a questo link: https://www.equivant.com/response-to-propublica-demonstrating-accuracy-equity-and-predictive-parity/(20 maggio 2022).

sarebbero state abbastanza accurate per essere usate nel suo caso. Dopo essere stato prima accolto in primo grado[527], poi respinto in appello[528], il ricorso è giunto fino alla Corte Suprema canadese, che ha riconosciuto le ragioni di Ewert. In particolare, la *majority opinion* a cura di Justice Wagner ha evidenziato come il CSC non abbia mai svolto verifiche sull'affidabilità dei sistemi di *riskassesment*coinvolti, quando siano utilizzati per decisioni riguardanti membri di minoranze etniche e culturali, anche dopo che erano state evidenziate preoccupazioni specifiche sul punto[529]. L'*opinion*, inoltre, sottolinea l'esistenza di evidenze statistiche di un trattamento deteriore degli imputati appartenenti a popoli indigeni da parte del sistema giudiziario canadese, i quali, mediamente, sono condannati a pene più lunghe e ricevono prognosi di rischio peggiori. L'utilizzo sempre più frequente di tecnologie avanzate, si legge nella sentenza, presenta il rischio di acuire queste discriminazioni e rendere la loro identificazione ancor più difficile di quanto non sia oggi, nascondendole dietro un velo di presunta oggettività tecnico-scientifica [530] . Per questo, l'introduzione nel procedimento penale di strumenti di *risk assesment* come quelli contestati da Ewert dev'essere accompagnata da verifiche particolarmente rigorose, che la Corte Suprema canadese, nel caso di specie, ha ritenuto mancare.

Sia la Corte del Wisconsin che quella canadese, dunque, riconoscono che il coinvolgimento di tecnologie avanzate, in decisioni delicate come quelle sulla libertà personale, solleva questioni urgenti: la sentenza sul caso *Loomis* evidenzia come l'algoritmo non possa mai generare una decisione realmente individualizzata sul singolo imputato, quella sul caso *Ewert* pone l'accento sul rischio di discriminazioni che si annida dietro questi strumenti. In ambo i casi, inoltre, la soluzione di queste criticità è individuata, in primo luogo, nel giudice, a cui è attribuito il compito di mantenere il pieno controllo sulla decisione, dando il giusto peso all'indicazione della tecnologia. Da questo punto di vista, però, la differenza tra le due pronunce è radicale. La Corte del Wisconsin, infatti, pare fidarsi del giudice quasi ciecamente – e in modo forse eccessivo – considerandolo capace di valutare correttamente anche i risultati di un sistema del quale, in ultima analisi, non

[527] Federal Court (Phelan J.), 2015 FC 1093, 343 C.R.R. (2d) 15.
[528]Federal Court of Appeal (Dawson J.A., Nadon and Webb JJ.A. Concurring), 2016 FCA 203, 487 N.R. 107.
[529]*Ewert v. Canada*, 2018 SCC 30 [2018] 2 S.C.R. 165 par. 46-67.
[530] «The clear danger posed by the CSC's continued use of assessment tools that may overestimate the risk posed by Indigenous inmates is that it could unjustifiably contribute to disparities in correctional outcomes in areas in which Indigenous offenders are already disadvantaged. For example, if the impugned tools overestimate the risk posed by Indigenous inmates, such inmates may experience unnecessarily harsh conditions while serving their sentences, including custody in higher security settings and unnecessary denial of parole. Overestimation of the risk may also contribute to reduced access to rehabilitative opportunities, such as a loss of the opportunity to benefit from a gradual and structured release into the community on parole before the expiry of a fixed-term sentence. Another effect of an overestimation of the risk is that it could bar an inmate from participation in Indigenous-specific programming that is contingent on an offender having a low security classification or being eligible for an escorted temporary absence […]In the context of the case at bar, this required, at the very least, that the CSC take seriously the credible concerns that have been repeatedly raised according to which information derived from the impugned tools is of questionable validity with respect to Indigenous inmates because the tools fail to account for cultural differences» *Ewert v. Canada*, par. 65-66; cfr. anche par. 57 ss.

conosce in modo chiaro il funzionamento, anche per motivi di segreto industriale[531]. La Corte Suprema canadese, invece, pare molto più consapevole delle difficoltà che la relazione uomo-macchina pone quando si inserisce nei meccanismi decisionali e dei potenziali pregiudizi insiti nella valutazione del giudice, testimoniati dal trattamento diseguale ricevuto nel sistema penale dai membri di alcune minoranze. Ne deriva la necessità, agli occhi dei giudici canadesi, di unattento esame della qualità del contributo della tecnologia in questo genere di valutazioni, poiché esso deve rappresentare uno strumento di mitigazione di un *bias* potenzialmente discriminatorio che pare presente nelle decisioni dei giudici umani[532]. Non è ragionevole, quindi, ipotizzare che avvenga l'esatto contrario, e che siano quest'ultimi a disinnescare i possibili pregiudizi insiti nelle indicazioni dei sistemi di supporto alla decisione di cui si avvalgono.

Come già detto, i casi *Loomis* ed *Ewert* sembrano, allo stato dell'arte, i più significativi tra i pochi in cui l'utilizzo di sistemi di intelligenza artificiale nei meccanismi decisionali è stato discusso di fronte a un giudice. Infatti, la diffusione di strumenti come COMPAS nei sistemi penali nordamericani, anche in decisioni che riguardano la libertà personale, non ha paragoni nelle altre democrazie mature. In Europa, in particolare, l'ipotesi di utilizzare strumenti di tal genere nelle valutazioni dei giudici è confinata al piano teorico e suscita profonde discussioni, e tecnologie

[531] Particolarmente significativo, da questo punto di vista, pare l'elenco di avvertimenti ai giudici che la sentenza *Loomis*prevede che accompagni ogni PSI, che dà un quadro efficace della difficoltà della valutazione cui questi ultimi sono chiamati al momento di valorizzare l'indicazione algoritmica nella decisione giudiziale (p. 769): «a circuit court must explain the factors in addition to a COMPAS risk assessment that independently support the sentence imposed. A COMPAS risk assessment is only one of many factors that may be considered and weighed at sentencing. Any Presentence Investigation Report ("PSI") containing a COMPAS risk assessment filed with the court must contain a written advisement listing the limitations.Additionally, this written advisement should inform sentencing courts of the following cautions as discussed throughout this opinion: The proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are determined. • Because COMPAS risk assessment scores are based on group data, they are able to identify groups of high-risk offenders — not a particular high-risk individual. • Some studies of COMPAS risk assessment scores have raised questions about whether they disproportionately classify minority offenders as having a higher risk of recidivism. • A COMPAS risk assessment compares defendants to a national sample, but no cross-validation study for a Wisconsin population has yet been completed. Risk assessment tools must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations. • COMPAS was not developed for use at sentencing, but was intended for use by the Department of Corrections in making determinations regarding treatment, supervision, and parole».

[532] In particolare, la perdurantemancanza di verificheapprofondite e risolutivedeidubbi su eventuali*bias* del sistema da parte del *Correctional Service of Canada* è considerato dallasentenza il principaleelemento a support dellepretese di Ewert, e rappresenta una chiaraviolazione del *Correction and Conditional Release Act* del 1992 (par. 86-87): «The fact that a review of the CSC's assessment tools was under way in 2005 was an important factor in the Federal Court's decision to dismiss Mr. Ewert's application for judicial review with respect to the resolution of his grievance: Ewert (2007), at paras. 66-67. It was also an important consideration in the Federal Court of Appeal's decision to uphold the dismissal of that application, including on the basis that it was premature: Ewert (2008), at paras. 7-8 and 10. In its 2007 decision, the Federal Court urged the CSC to explain to Mr. Ewert the results, if any, of its review. Such an explanation had not yet been provided when Mr. Ewert appealed to the Federal Court of Appeal in 2008 — eight years after he commenced the grievance procedure. Indeed, the trial judge in the present proceeding observed that there was no evidence that the CSC had ever completed the research referred to by the Federal Court in 2007 and anticipated by the Federal Court of Appeal in 2008: para. 72. Almost two decades have now passed since Mr. Ewert first complained about the use of certain of the impugned assessment tools with respect to Indigenous inmates. In these exceptional circumstances, Mr. Ewert should not be required to begin the grievance process anew in order to determine whether the CSC's continued failure to address the validity of the impugned assessment tools is a breach of its duty under s. 24(1) of the CCRA».

basate sull'intelligenza artificiale, nei sistemi penali, sono usate quasi esclusivamente nella fase delle indagini[533]. Ciò nonostante, anche nel continente europeo, e in particolare in Italia, le corti hanno avuto la possibilità di occuparsi delle conseguenze dell'uso di algoritmi nei meccanismi decisionali, sia da parte della Pubblica Amministrazione (è stato il caso, come si dirà, del sistema utilizzato per l'assegnazione ai docenti delle loro sedi di lavoro nella procedura concorsuale indetta con le riforme della c.d. *buona scuola*) che di poteri privati, con la conseguente applicazione delle più volte menzionate norme di diritto antidiscriminatorio (è la vicenda dell'algoritmo di assegnazione dei turni di lavoro dei *rider* della piattaforma *Deliveroo*).

Il caso *Deliveroo*nasceva da un ricorso al Giudice del Lavoro di tre distinte sezioni della CGIL di Bologna, che, avvalendosi della possibilità di agire in giudizio in quanto enti esponenziali particolarmente rappresentativi garantita dall'art. 5 c. 2 del D.Lgs. 216/2003[534] – per l'appunto, una delle principali norme di diritto antidiscriminatorio nel nostro ordinamento - accusavano l'algoritmo *Frank*, usato dalla piattaforma, di discriminazioni illecite nel trattamento del personale impiegato nelle consegne a domicilio. Al centro del problema stava l'ampia mole di dati utilizzati dal sistema per la profilazione dei *rider* e in particolare la circostanza che, dall'eventuale assenza da un turno di lavoro da essi scelto, non comunicata con almeno 24 ore di anticipo, derivasse un'importante penalizzazione da parte dell'algoritmo, che limitava la possibilità di prenotare nuove sessioni di lavoro. Da questo punto di vista, il sistema non consentiva alcuna distinzione riguardo le ragioni dell'assenza. Secondo i sindacati ricorrenti, ciò avrebbe scoraggiato i lavoratori dall'adesione a ogni possibile iniziativa di sciopero, posto che avrebbe portato, inevitabilmente, a una sanzione che li avrebbe messi ai margini dell'economia della piattaforma. Questo avrebbe rappresentato una

[533] Cfr. M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei riskassessmenttools tra Stati Uniti ed Europa*, in *Diritto Penale Contemporaneo*, 2019, accessedJune 21, 2022, https://bit.ly/3ObOWwH (20 maggio 2022); D. POLIDORO, *Tecnologie Informatiche e Procedimento Penale: La Giustizia Penale 'Messa Alla Prova' Dall'intelligenza Artificiale*, in *Archivio penale*, 3, 2020. Il sistema più avanzato utilizzato in Europa è, probabilmente, HART (acronimo di *Harm Assessment Risk Tool*), un sistema sperimentato dalla polizia della città di Durham, nel Regno Unito (e sospettato di esiti discriminatori verso le fasce più povere della popolazione) per la valutazione della pericolosità di individui sospettati di aver commesso determinati reati, al fine di supportare la decisione riguardo alla loro ammissione a programmi di riabilitazione che, nel sistema penale britannico, sono alternativi all'esercizio dell'azione penale. In letteratura cfr. 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 and Communications Technology Law*, 2018, p. 227 ss. Si rimanda, inoltre, all'analisi degli utilizzi dell'intelligenza artificiale nel sistema giustizia svolta *infra*, nell'ultima parte del lavoro, cfr. p. 256 ss.

[534] L'art. 5 del D.Lgs. 216/2003, *Attuazione della direttiva 2000/78/CE per la parita' di trattamento in materia di occupazione e di condizioni di lavoroe della direttiva n. 2014/54/UE relativa alle misure intese ad agevolare l'esercizio dei diritti conferiti ai lavoratori nel quadro della libera circolazione dei lavoratori* fornisce a sindacati e altri enti esponenziali legittimazione ad agire in giudizio su delega della vittima di discriminazione o nel caso di discriminazione collettiva. La norma recita: «1. Le organizzazioni sindacali, le associazioni e le organizzazioni rappresentative del diritto o dell'interesse leso, in forza di delega, rilasciata per atto pubblico o scrittura privata autenticata, a pena di nullita', sono legittimate ad agire ai sensi dell'articolo 4, in nome e per conto o a sostegno del soggetto passivo della discriminazione (e dei suoi familiari), contro la persona fisica o giuridica cui È riferibile il comportamento o l'atto discriminatorio. 2. I soggetti di cui al comma 1 sono altresì legittimati ad agire nei casi di discriminazione collettiva qualora non siano individuabili in modo diretto e immediato le persone lese dalla discriminazione».

discriminazione illecita tra di essi sulla base dell'adesione a iniziative sindacali, ricondotta da pacifica giurisprudenza nella categoria delle «convinzioni personali», uno dei fattori di discriminazione proibiti dal citato D.lgs. 216/2003[535]. Il Giudice del Lavoro di Bologna ha riconosciuto la bontà di tali pretese, riconoscendo la natura discriminatoria delle strategie di profilazione dei *rider* adottate da *Deliveroo*, condannando quest'ultima a risarcire le organizzazioni sindacali ricorrenti e a dare ampia pubblicità al provvedimento[536].

Come già anticipato, il tema della discriminazione algoritmica è menzionato anche da alcune sentenze emesse dal Consiglio di Stato riguardo a quello che è ormai noto come il caso c.d. *buona scuola*[537]. La vicenda ha origine dal piano straordinario di assunzioni nella scuola secondaria previsto con la Legge 107/2015 (la riforma detta, appunto, della *buona scuola*)[538] al fine di tamponare il problema, ormai sistemico, del lavoro precario nel sistema di pubblica istruzione. Parte delle procedure previste, e in particolare l'assegnazione agli insegnanti di determinate sedi di lavoro in funzione del collocamento in graduatoria e delle loro preferenze, era stata automatizzata, con esiti contrastanti. In molti casi, infatti, docenti posizionatisi nei primi posti sono stati assegnati a sedi di lavoro distanti dalla provincia di residenza, e in classi di concorso e ordini di scuole diversi da quanto indicato al momento dell'iscrizione. Allo stesso tempo, candidati posizionatisi in posizioni peggiori, assegnati alla rispettiva sede di lavoro solo in una fase successiva ai migliori, avevano finito per ottenere classe di concorso, scuola e sede geograficadi preferenza, rimaste vacanti al termine della prima fase. I ricorsi amministrativi che ne sono derivati, prima al T.A.R. del Lazio territorialmente competente, poi al Consiglio di Stato quale giudice di secondo grado, hanno portato a una serie di pronunce contenenti considerazioni estremamente significative sul ruolo degli algoritmi nelle decisioni della Pubblica Amministrazione e sulle strategie giuridiche, tradizionali e

[535] È l'art. 2 c.1 del citato D.Lgs 261/2003 a definire i fattori di discriminazione proibiti: «Ai fini del presente decreto e salvo quanto disposto dall'articolo 3, commi da 3 a 6, per principio di parità di trattamento si intende l'assenza di qualsiasi discriminazione diretta o indiretta a causa della religione,delle convinzioni personali, degli handicap, dell'età, della nazionalità o dell'orientamento sessuale». A chiarire che gli orientamenti in materia sindacale rientrano tra le condizioni personali protette dalla norma è stata di recente la stessa Suprema Corte, cfr. Cass. civ. sez. lav., 2 gennaio 2020, n. 1, richiamata anche dal Tribunale di Bologna.

[536] Tribunale ordinario di Bologna – sez. lavoro, ord. 31 dicembre 2020, in *Bollettino Adapt*, 2021, https://bit.ly/3NmRLtL (21 maggio 2022). In dottrina cfr. S. BORELLI, M. RANIERI, *La discriminazione nel lavoro autonomo. Riflessioni a partiredall'algoritmo Frank*, in*Labour& Law Issues*, 2021, http://labourlaw.unibo.it/article/view/13169 (21 giugno 2022); M. MARRONE, *Rights against the machine! Food delivery, piattaforme digitali e sindacalismo informale*, in *Labour& Law Issues*, 5, 1, 2019;M. BORZAGA, M. MAZZETTI,*Discriminazioni algoritmiche e tutela dei lavoratori: riflessioni a partire dall'Ordinanza del Tribunale di Bologna del 31 dicembre 2020*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2022, p. 225-250; M. FASCIGLIONE, *Gig economy e diritti fondamentali sul lavoro in una recente sentenza del Tribunale di Bologna*, in*IRiSS*, 9 febbraio 2021, https://bit.ly/3OuXm1M (21 maggio 2022).

[537] Si tratta, in particolare, di Cons. St. sez. VI, 13 dicembre 2019, n. 8474 e Cons. St. sez. VI, 4 febbraio 2020, n. 881.

[538] Legge n. 107 del 13 luglio 2015, *Riforma del sistema nazionale di istruzione e formazione e delega per il riordino delle disposizioni legislative vigenti*.

innovative, per farvi fronte[539]. Si tratta di sentenze che verranno prese in considerazione più approfonditamente nella terza parte di questo lavoro, dedicata, per l'appunto, ai possibili *nuovi diritti* che l'era dell'intelligenza artificiale impone di riconoscere all'individuo. Per quanto riguarda la discriminazione algoritmica, rileva in particolare quanto espresso, peraltro con formulazione pressochè identica, in due sentenze del Consiglio di Stato, la n. 8474 del 2019 e la n. 881 del 2020. Entrambe le pronunce rappresentano, come vedremo, una significativa apertura all'utilizzo di tecnologie avanzate, anche basate sull'intelligenza artificiale, nel procedimento amministrativo, purché siano presenti determinate garanzie (considerate assenti nel caso di specie, ragion per cui ambo le sentenze riconoscono l'illegittimità di quanto avvenuto nella vicenda *buona scuola*). Tra queste, è presente anche il rispetto di un «principio di non discriminazione algoritmica», che le due pronunce agganciano al Cons. n. 71 del GDPR, una delle disposizioni più significative su questo tema, come si dirà al paragrafo successivo[540]. Di particolare interesse, inoltre, appare quanto

[539] Oltre alla già citata Cons. St. VI, 4 febbraio 2020, n. 881, sono state estensivamente commentate la pronuncia del T.A.R. Lazio-Roma sez. IIIbis, 10 settembre 2018, n. 9227, poi confermata in appello dalla menzionata Cons. St. sez. VI, 13 dicembre 2019, n. 8474, pur con argomentazioni di molto differenti riguardo all'ammissibilità dell'uso di algoritmi nei procedimenti amministrativi, e la sentenza Cons. St. 8 aprile 2019, n. 2270, analizzate nel dettaglio *infra,* nella terza parte di questo lavoro. Per dei commenti in dottrina cfr. I. A. NICOTRA, V. VARONE, *L'algoritmo, intelligente ma non troppo*, in *Rivista AIC*, 4, 2019, p. 86-106; E. COCCHIARA, *Procedimento amministrativo e "buon algoritmo"*, in *amministrativ@mente*, 3, 2020, p. 370-385; L. MUSSELLI, *La decisione amministrativa nell'età degli algoritmi: primi spunti*, in *MediaLaws –Rivista di diritto dei media*, 1, 2020, p. 18-28; A. SIMONCINI, *L'algoritmo incostituzionale: l'intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019, p. 63 ss.

[540] L'art. 71 del GDPR, come si vedrà nel dettaglio più avanti, detta le linee generali della regolazione della decisione automatizzata sull'individuo, e delle garanzie minime in capo a quest'ultimo in materia di controllo e revisione umana. Tali orientamenti sono tradotti in disciplina vincolante dal successivo art. 22. Il testo del Considerando 71 è: «L'interessato dovrebbe avere il diritto di non essere sottoposto a una decisione, che possa includere una misura, che valuti aspetti personali che lo riguardano, che sia basata unicamente su un trattamento automatizzato e che produca effetti giuridici che lo riguardano o incida in modo analogo significativamente sulla sua persona, quali il rifiuto automatico di una domanda di credito online o pratiche di assunzione elettronica senza interventi umani. Tale trattamento comprende la «profilazione», che consiste in una forma di trattamento automatizzato dei dati personali che valuta aspetti personali concernenti una persona fisica, in particolare al fine di analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze o gli interessi personali, l'affidabilità o il comportamento, l'ubicazione o gli spostamenti dell'interessato, ove ciò produca effetti giuridici che la riguardano o incida in modo analogo significativamente sulla sua persona. Tuttavia, è opportuno che sia consentito adottare decisioni personali sulla base di tale trattamento, compresa la profilazione, se ciò è espressamente previsto dal diritto dell'Unione o degli Stati membri cui è soggetto il titolare del trattamento, anche a fini di monitoraggio e prevenzione delle frodi e dell'evasione fiscale secondo i regolamenti, le norme e le raccomandazioni delle istituzioni dell'Unione o degli organismi nazionali di vigilanza e a garanzia della sicurezza e dell'affidabilità di un servizio fornito dal titolare del trattamento, o se è necessario per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento, o se l'interessato ha espresso il proprio consenso esplicito. In ogni caso, tale trattamento dovrebbe essere subordinato a garanzie adeguate, che dovrebbero comprendere la specifica informazione all'interessato e il diritto di ottenere l'intervento umano, di esprimere la propria opinione, di ottenere una spiegazione della decisione conseguita dopo tale valutazione e di contestare la decisione. Tale misura non dovrebbe riguardare un minore. Al fine di garantire un trattamento corretto e trasparente nel rispetto dell'interessato, tenendo in considerazione le circostanze e il contesto specifici in cui i dati personali sono trattati, è opportuno che il titolare del trattamento utilizzi procedure matematiche o statistiche appropriate per la profilazione, metta in atto misure tecniche e organizzative adeguate al fine di garantire, in particolare, che siano rettificati i fattori che comportano inesattezze dei dati e sia minimizzato il rischio di errori e al fine di garantire la sicurezza dei dati personali secondo una modalità che tenga conto dei potenziali rischi esistenti per gli interessi e i diritti dell'interessato e impedisca, tra l'altro, effetti discriminatori nei confronti di persone fisiche sulla base della razza o dell'origine etnica, delle opinioni politiche, della religione o delle convinzioni personali, dell'appartenenza sindacale, dello status genetico, dello stato di salute o dell'orientamento sessuale, ovvero un trattamento che comporti misure aventi tali effetti. Il processo decisionale automatizzato e la

affermato da ambo le pronunce, con le stesse parole, in materia di percezione della tecnologia, formazione dei *dataset* e *design* dell'algoritmo: «In molti campi gli algoritmi promettono di diventare lo strumento attraverso il quale coreggere le storture e le imperfezioni che caratterizzano tipicamente i processi cognitivi e le scelte compiute dagli esseri umani, messe in luce soprattutto negli ultimi anni da un'imponente letteratura di economia comportamentale e psicologia cognitiva. In tale contesto, le decisioni prese dall'algoritmo assumono così un'aura di neutralità, frutto di asettici calcoli razionali basati su dati. Peraltro, già in tale ottica è emersa altresì una lettura critica del fenomeno, in quanto l'impiego di tali strumenti comporta in realtà una serie di scelte e di assunzioni tutt'altro che neutre: l'adozione di modelli predittivi e di criteri in base ai quali i dati sono raccolti, selezionati, sistematizzati, ordinati e messi insieme, la loro interpretazione e la conseguente formulazione di giudizi sono tutte operazioni frutto di precise scelte e di valori, consapevoli o inconsapevoli; da ciò ne consegue che tali strumenti sono chiamati ad operare una serie di scelte, le quali dipendono in gran parte dai criteri utilizzati e dai dati di riferimento utilizzati, in merito ai quali è apparso spesso difficile ottenere la necessaria trasparenza»[541]. Si tratta com'è evidente, di considerazioni particolarmente significative, che dimostrano un elevato grado di consapevolezza, da parte del massimo organo della Giustizia Amministrativa italiana, della complessità del problema della discriminazione algoritmica e delle sue molte sfaccettature. Una profondità d'analisi che supera di gran lunga quella dimostrata dai provvedimenti che hanno deciso i visti casi *Loomis* ed *Ewert*, e fa sì che le sentenze del Consiglio di Stato in esame siano, probabilmente, tra gli approdi giurisprudenziali in cui il tema è affrontato nel modo più completo e maturo sullo scenario globale. Ciò nonostante, è del tutto evidente che vicende simili a quelle riassunte siano destinate a divenire sempre più comuni, ed essere, di conseguenza, sottoposte sempre più di frequente all'attenzione delle corti. Com'è evidente che il tema della discriminazione algoritmica rappresenti una sfida anche per i legislatori, per quanto le capacità di elaborare soluzioni particolarmente innovative da parte della giurisprudenza non siano da sottovalutare, anche negli ordinamenti di *civil law*. Alle ipotesi di regolazione più significative del fenomeno è, per l'appunto, dedicato il prossimo paragrafo.

---

profilazione basati su categorie particolari di dati personali dovrebbero essere consentiti solo a determinate condizioni». Com'è noto, i *Considerando* fanno parte del Preambolo degli atti giuridici dell'Unione e ne rappresentano la motivazione, non hanno valore precettivo e non contengono dichiarazioni di natura politica, cfr. *Manuale interistituzionale di convenzioni redazionali*, Ufficio delle Pubblicazioni dell'Unione Europea, 2022, https://publications.europa.eu/code/it/it-120200.htm (21 maggio 2022). Essi, comunque, nel corso del tempo hanno acquisito un importante valore interpretativo (tanto da rappresentare, talvolta, uno strumento argomentativo per le giurisdizioni superiori degli stati membri, come nelle sentenze qui in esame).

[541] Cfr. Cons. St. sez. VI, 13 dicembre 2019, n. 8474 par. 7.1-7.2 e Cons. St. VI, 4 febbraio 2020, n. 881 par. 5.1-5.2.

6. Costruire un diritto dell'intelligenza artificiale a prova di discriminazione: lo stato dell'arte e le possibili prospettive *de iure condendo*

A prima vista, la strategia più immediata ed efficace per contrastare la discriminazione algoritmica potrebbe sembrare l'emanazione di nuove norme a tutela dell'eguaglianza, che aggiornino il *corpus* del diritto antidiscriminatorio. Questa, però, probabilmente si rivelerebbe una strategia inefficace: il problema centrale sollevato dall'implementazione di tecnologie basate sull'intelligenza artificiale nei meccanismi decisionali sta nell'occultamento di discriminazioni tradizionali, prima che nella creazione di discriminazioni di nuovo genere. Come già visto, infatti, a prescindere da un eventuale aggiornamento del loro catalogo, i fattori di discriminazione proibiti rischierebbero comunque di ripresentarsi sotto mentite spoglie, attraverso l'elaborazione di *proxies* apparentemente neutri. L'esperienza concreta – si pensi ai casi *Ewert* e *Loomis* – ha già mostrato come le questioni più spinose sorgano dalla difficoltà di identificare discriminazioni chiaramente vietate dalla legge, e non da vuoti legislativi nelle norme a tutela dell'eguaglianza. Quanto detto non esclude, in ogni caso, che la necessità di aggiornare il *corpus* di norme di diritto antidiscriminatorio potrebbe comunque presentarsi, com'è già avvenuto più volte, in parallelo a mutamenti socio-culturali. Né può escludersi che ciò avvenga anche a causa del sempre maggiore utilizzo di algoritmi nei procedimenti decisionali o valutativi, che potrebbe generare tali mutamenti socio-culturali, oggi difficilmente prevedibili.

Pare, dunque, che la regolazione della discriminazione algoritmica debba percorrere altre strade. In particolare, una regolazione efficace dovrebbe orentarsi verso due principali obiettivi: minimizzare la presenza di *bias*negli algoritmi e massimizzare la possibilità di identificare e correggere eventuali indicazioni discriminatorie. Peraltro, allo stato dell'arte quasi non si rinvengono testi normativi che, nell'intento di regolare le nuove tecnologie,menzionino il tema dei possibili esiti discriminatori di alcuni loro utilizzi. Le poche iniziative esistenti provengono principalmente dall'Unione Europea, a conferma del ruolo di guida sullo scenario globale che quest'ultima intende assumere in materia di "diritti digitali"[542].

[542] Infatti, volgendo lo sguardo ad alcune delle principali esperienze democratiche extraeuropee, e prendendo ad esempio la normativa in materia di protezione dei dati personali, la più idonea a condurre tal genere di analisi, posta l'assenza, allo stato dell'arte, di discipline autonome delle tecnologie basate sull'intelligenza artificiale, non può non notarsi che, banalmente, la parola *discrimination* non compare mai nel *Data Protection Act* britannico del 2018, nel *Data Availability and Transparency Act* e nella *Privacy Regulation* del 2013 australiani, né nel *Personal Information Protection and Electronic Documents Act* canadese del 2000. La legge giapponese sulla *privacy*, invece, utilizza il termine una volta sola, al momento di definire la categoria dei dati sensibili (una traduzione non ufficiale della norma è disponibile a questo link: https://bit.ly/3HDGcwZ - 21 maggio 2022). Per quanto ciascuno di questi testi normativi preveda tutele almeno parzialmente accostabili a quelle previste dalla legislazione di matrice eurounitaria, che si ripercuotono certamente anche sul rischio di discriminazione algoritmica, ad esempio in materia di trasparenza di natura e finalità del trattamento o di accuratezza e attualità dei dati, non può non evidenziarsi, anche dal punto di vista simbolico, che la parola *discrimination* appare quattro volte nella versione inglese del GDPR, e ben 26 nella Proposta di Regolamento in materia di intelligenza artificiale del 2021. Nonostante l'insieme delle tutele da essa previste abbia

A prendere in considerazione l'utilizzo della tecnologia nei procedimenti valutativi è prima di tutto il GDPR, che, all'art. 22 e al Considerando 71, fa riferimento al «processo decisionale automatizzato». Il Regolamento detta alcuni accorgimenti volti ad assicurare che il contesto in cui si sviluppa la decisione algoritmica garantisca sempre un determinato livello di controllo umano, anche al fine di evitare risultati discriminatori. L'art. 22 codifica un vero e proprio diritto individuale a non essere sottoposto a una decisione totalmente automatizzata, pur con importantissime eccezioni: è sufficiente, infatti, che la decisione totalmente algoritmica si basi sul consenso dell'interessato al trattamento, sia necessaria per l'esecuzione di un contratto tra egli e il titolare o sia autorizzata dalla legge[543]. Anche in tali casi, però, all'individuo oggetto della decisione dev'essere garantita la possibilità di richiedere l'intervento umano, esprimere la propria opinione e contestare la decisione. Egli, inoltre, dev'essere sempre informato sulla tecnologia utilizzata dal sistema impiegato nel procedimento decisionale. Si tratta di accorgimenti – possibilità di controllo umano e spiegazione del funzionamento dei sistemi – che trascendono il tema della discriminazione algoritmica, venendo in gioco nell'analisi di pressochè ogni applicazione dell'intelligenza artificiale che abbia un impatto sociale, con sfaccettature che verranno indagate in profondità nella terza parte del lavoro. Completa il quadro di quanto previsto dal GDPR il menzionato Considerando 71, che, al secondo paragrafo, incarica il titolare del trattamento corrispondente a un processo decisionale automatizzato di «mettere in atto misure tecniche e organizzative adeguate al fine di garantire, inparticolare, che siano rettificati i fattori che comportano inesattezze dei dati e sia minimizzato ilrischio di errori, e al fine di garantire la sicurezza dei dati personali secondo una modalità che

---

sicuramente un effetto deterrente riguardo a possibili esiti discriminatori, il termine non compare mai nemmeno nella vista *Directive on automated decision making* canadese del 1 aprile 2019. Inoltre, l'attenzione dell'Unione Europea per l'intreccio tra diritti fondamentali e nuove tecnologie è ben rappresentata dalla proposta, da parte della Commissione Europea, di una *Dichiarazione europea sui diritti e i principi digitali per il decennio digitale*, COM(2022) 28 final. In letteratura cfr. P. DE PASQUALE, *Verso una Carta dei diritti digitali (fondamentali) dell'Unione Europea?*, in *Il diritto dell'Unione Europea*, 3, 2022. V. anche O. POLLICINO, *L' "autunno caldo" della Corte di giustizia in tema di tutela dei diritti fondamentali in rete e le sfide del costituzionalismo alle prese con i nuovi poteri privati in ambito digitale*, in *Federalismo.it*, 19, 2019.

[543] L'art. 22 GDPR recita: «L'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona. 2. Il paragrafo 1 non si applica nel caso in cui la decisione: a) sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento; b) sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato; c) si basi sul consenso esplicito dell'interessato. 3. Nei casi di cui al paragrafo 2, lettere a) e c), il titolare del trattamento attua misure appropriate per tutelare i diritti, le libertà e i legittimi interessi dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione. 4. Le decisioni di cui al paragrafo 2 non si basano sulle categorie particolari di dati personali di cui all'articolo 9, paragrafo 1, a meno che non sia d'applicazione l'articolo 9, paragrafo 2, lettere a) o g), e non siano in vigore misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato». Per dei commenti v. A. ODDENINO, *Decisioni algoritmiche e prospettive internazionali di valorizzazione dell'intervento umano*, in *DPCE online*, 1, 2020, p. 199 ss; A. CAIA, *Art. 22*, in G. M. RICCIO, G. SCORZA, E. BELISARIO (a cura di), *GDPR e normativa privacy. Commentario*, 2018, p. 219 ss. Criticano l'ampiezza delle eccezioni previste dalla norma C. CASONATO, *Costituzione e intelligenza artificiale cit.*, p. 723-724; A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019, p. 79 ss.

tengaconto dei potenziali rischi esistenti per gli interessi e i diritti dell'interessato e impedisca, tra l'altro,effetti discriminatori nei confronti di persone fisiche sulla base della razza o dell'origine etnica,delle opinioni politiche, della religione o delle convinzioni personali, dell'appartenenza sindacale,dello status genetico, dello stato di salute o dell'orientamento sessuale, ovvero un trattamento che comporti misure aventi tali effetti»[544]. Al netto del suo valore unicamente interpretativo, da quest'ultima disposizione emerge la volontà di minimizzare, con adeguate misure tecniche, la presenza di *bias* che possano condurre a risultati discriminatori, una finalità che sembra presente anche nella Proposta di Regolamento in materia di intelligenza artificiale.

Il progetto normativo presentato dalla Commissione Europea il 21 aprile 2021, infatti, menziona in vari punti il rischio che alcune applicazioni dell'intelligenza artificiale conducano ad esiti discriminatori. Il rimedio è identificato in modo deciso nello sviluppo di pratiche avanzate di *governance*dei dati, idonee a garantire *dataset* di alta qualità, decisivi, com'è noto, per il corretto funzionamento di molti sistemi di intelligenza artificiale. L'obiettivo è ottenere set di dati «pertinenti, rappresentativi, esenti da errori e completi»[545] e per raggiungerlo sono previste misure molto rigorose, verso le quali non sono mancate critiche da parte di alcuni addetti ai lavori, che le considerano eccessivamente onerose[546]. In particolare, dovranno svilupparsi strategie di gestione dei

[544] Per il testo completo della norma v. *supra*, n. 540.

[545] Sono le parole dell'art. 10 (*Dati e governance dei dati*) della Proposta. La norma, nella sua interezza, recita: «1. I sistemi di IA ad alto rischio che utilizzano tecniche che prevedono l'uso di dati per l'addestramento di modelli sono sviluppati sulla base di set di dati di addestramento, convalida e prova che soddisfano i criteri di qualità di cui ai paragrafi da 2 a 5. 2. I set di dati di addestramento, convalida e prova sono soggetti ad adeguate pratiche di governance e gestione dei dati. Tali pratiche riguardano in particolare: a) le scelte progettuali pertinenti; b) la raccolta dei dati; c) le operazioni di trattamento pertinenti ai fini della preparazione dei dati, quali annotazione, etichettatura, pulizia, arricchimento e aggregazione; d) la formulazione di ipotesi pertinenti, in particolare per quanto riguarda le informazioni che si presume che i dati misurino e rappresentino; e) una valutazione preliminare della disponibilità, della quantità e dell'adeguatezza dei set di dati necessari; f) un esame atto a valutare le possibili distorsioni; g) l'individuazione di eventuali lacune o carenze nei dati e il modo in cui tali lacune e carenze possono essere colmate. 3. I set di dati di addestramento, convalida e prova devono essere pertinenti, rappresentativi, esenti da errori e completi. Essi possiedono le proprietà statistiche appropriate, anche, ove applicabile, per quanto riguarda le persone o i gruppi di persone sui quali il sistema di IA ad alto rischio è destinato a essere usato. Queste caratteristiche dei set di dati possono essere soddisfatte a livello di singoli set di dati o di una combinazione degli stessi. 4. I set di dati di addestramento, convalida e prova tengono conto, nella misura necessaria per la finalità prevista, delle caratteristiche o degli elementi particolari dello specifico contesto geografico, comportamentale o funzionale all'interno del quale il sistema di IA ad alto rischio è destinato a essere usato. 5. Nella misura in cui ciò sia strettamente necessario al fine di garantire il monitoraggio, il rilevamento e la correzione delle distorsioni in relazione ai sistemi di IA ad alto rischio, i fornitori di tali sistemi possono trattare categorie particolari di dati personali di cui all'articolo 9, paragrafo 1, del regolamento (UE) 2016/679, all'articolo 10 della direttiva (UE) 2016/680 e all'articolo 10, paragrafo 1, del regolamento (UE) 2018/1725, fatte salve le tutele adeguate per i diritti e le libertà fondamentali delle persone fisiche, comprese le limitazioni tecniche all'utilizzo e al riutilizzo delle misure più avanzate di sicurezza e di tutela della vita privata, quali la pseudonimizzazione o la cifratura, qualora l'anonimizzazione possa incidere significativamente sulla finalità perseguita. 6. Per lo sviluppo di sistemi di IA ad alto rischio diversi da quelli che utilizzano tecniche che prevedono l'addestramento di modelli si applicano adeguate pratiche di gestione e governance dei dati, al fine di garantire che tali sistemi di IA ad alto rischio siano conformi al paragrafo 2».

[546] Cfr. ad esempio P. GLAUNER, *An Assessment of the AI Regulation Proposed by the European Commission*, 26 maggio 2021, http://arxiv.org/abs/2105.15133 (20 maggio 2022), le dichiarazioni critiche raccolte in M. HEIKKILÄ, *A Quick Guide to the Most Important AI Law You've Never Heard of*, in *MIT Technology Review*, https://www.technologyreview.com/2022/05/13/1052223/guide-ai-act-europe/ (20 maggio 2022), o le osservazioni sulla

dati atte ad assicurare, tra le altre cose, che i *dataset* siano completi, idonei al raggiungimento degli obiettivi per essi previsti, e rappresentativi, ovvero possiedano le proprietà statistiche appropriate a rappresentare le specificità della popolazione e del contesto di riferimento. Dovranno essere previsti, inoltre, esami atti a identificare precocemente possibili lacune o carenze nei dati e distorsioni negli output, oltre alle modalità di pronta risoluzione di tali difetti[547]. Parallelamente, la Proposta di Regolamento ipotizza lo sviluppo di piattaforme di condivisione tra imprese, università e centri di ricerca di tali *dataset* di qualità, al fine di creare, come già dichiarato in precedenti documenti delle istituzioni, «spazi europei dei dati» che orientino e incentivino l'innovazione[548].

Dunque, sia il GDPR che la Proposta di Regolamento sull'intelligenza artificiale pongono alcuni primi presidi contro la discriminazione algoritmica. Se il Regolamento in materia di protezione dei dati personali pone l'enfasi sul ruolo dell'essere umano, tentando di assicurare che esso sia posto nelle condizioni più idonee per identificare e disinnescare eventuali discriminazioni generate dalla tecnologia, la Proposta in materia di intelligenza artificiale prende una strada diversa, puntando a minimizzare le possibilità di esiti distorti attraverso una rete di articolate norme tecniche[549]. Si tratta senza dubbio di iniziative promettenti, conformi a quelle che appaiono le strategie più idonee, allo

---

difficoltà di soddisfare taluni dei requisiti previsti dalla Proposta di Regolamento, pur nel contesto di una valutazione positiva dell'iniziativa legislativa, avanzate da DIGITALEUROPE, *Report – Digital Europe's Initial Findings on the Proposed AI Act*, Bruxelles, 6 agosto 2021, p. 4 ss. https://bit.ly/3xJvyQv (20 maggio 2022). Per una critica riguardante il possibile impatto economico della regolazione proposta, cfr. B. MUELLER, *How Much Will the Artificial Intelligence ActCost Europe?*, Report - Center for Data Innovation, 26 luglio 2021, https://bit.ly/3tRXCjs (20 maggio 2022); i cui risultati sono stati contestati da CENTER FOR EUROPEAN POLICY STUDIES,*Report - Clarifying the Costs for the EU's AI Act*, Bruxelles, 24 settembre 2021, https://bit.ly/3QGqa9o (20 maggio 2022). Non si può non rilevare, inoltre, come le scelte lessicali della Proposta di Regolamento si prestino a spinose ambiguità interpretative, che in caso di approvazione della corrente formulazione starà prima di tutto a operatori pratici e regolatori di settore risolvere: come definire, ad esempio, quando un *dataset* possa dirsi pienamente "rappresentativo"? Che parametro prendere a riferimento per tale giudizio?

[547] Cfr. gli artt. 10 ss. della Proposta.

[548] La finalità è dichiarata al Considerando 45 della Proposta: «Ai fini dello sviluppo di sistemi di IA ad alto rischio, è opportuno concedere ad alcuni soggetti, come fornitori, organismi notificati e altre entità pertinenti, quali i poli dell'innovazione digitale, le strutture di prova e sperimentazione e i ricercatori, l'accesso a set di dati di elevata qualità e la possibilità di utilizzarli nell'ambito dei rispettivi settori di attività connessi al presente regolamento. Gli spazi comuni europei di dati istituiti dalla Commissione e l'agevolazione della condivisione dei dati tra imprese e con i governi, nell'interesse pubblico, saranno fondamentali per fornire un accesso affidabile, responsabile e non discriminatorio a dati di elevata qualità a fini di addestramento, convalida e prova dei sistemi di IA. Ad esempio, per quanto riguarda la salute, lo spazio europeo di dati sanitari agevolerà l'accesso non discriminatorio ai dati sanitari e l'addestramento di algoritmi di intelligenza artificiale su tali set di dati in modo sicuro, tempestivo, trasparente, affidabile e tale da tutelare la vita privata, nonché con un'adeguata governance istituzionale. Le autorità competenti interessate, comprese quelle settoriali, che forniscono o sostengono l'accesso ai dati, possono anche sostenere la fornitura di dati di alta qualità a fini di addestramento, convalida e prova dei sistemi di IA». Le norme in materia di condivisione dei dati sono previste al Titolo V (*Misure a sostegno dell'innovazione*) della Proposta (artt. 53-55). La creazione di uno spazio europeo dei dati sanitari è al centro di un'ulteriore Proposta di Regolamento, presentata dalla Commissione al Parlamento Europeo e al Consiglio il 3 maggio 2022, cfr. COMMISSIONE EUROPEA, *Proposta di Regolamento del Parlamento Europeo e del Consiglio sullo spazio europeo dei dati sanitari*, COM(2022) 197 final.

[549] Si tratta, peraltro, di norme almeno in parte accostabili a quelle previste dalla *Directive on automated decision-making* canadese, che, come visto, pur in modo molto più generico rispetto alla Proposta europea, circonda di garanzie il coinvolgimento di tecnologie avanzate nelle decisioni della pubblica amministrazione, prevedendo adempimenti tecnici come lo svolgimento di un *algorithmic impact assessment* e di controlli di qualità sul corretto funzionamento del sistema, riguardanti anche la qualità dei dati utilizzati, cfr. *supra*, p. 70 ss.

stato dell'arte, per ridurre al minimo l'impatto del problema. D'altro canto, non si può non ribadire come la loro concreta realizzabilità tecnica susciti discussioni e perplessità, sia tra gli esperti di regolazione che di tecnologia[550]. La sostenibilità tecnologica ed economica della normativa di settore è un tema che accompagna ogni ipotesi di regolazione dell'intelligenza artificiale, come visto nei capitoli precedenti e come si dirà diffusamente nella terza parte di questo lavoro. Ciò nonostante, un dato non può essere trascurato: regole come quelle appena prese in considerazione – volte a migliorare le capacità d'azione del decisore umano, o ad assicurare *dataset*di qualità per lo sviluppo tecnologico – hanno di certo la capacità concreta di minimizzare il rischio connesso all'applicazione tecnologica di volta in volta considerata, in questo caso la discriminazione algoritmica. Questo a prescindere dalla loro concreta realizzabilità tecnica, questione che si risolve, in ultima analisi, nella definizione di uno standard di conformità (oltre il quale, ad esempio, un *dataset*è definibile "di qualità") in grado di tutelare i diritti cui la norma è posta a presidio senza frustrare il mercato e l'innovazione tecnologica. Si tratta di un bilanciamento di interessi estremamente complesso e variabile, ma non maggiore di quanto avviene con l'applicazione di norme volte a regolare altri settori dell'ordinamento ad elevato tasso di specificità tecnico-scientifica (si pensi alla tutela dell'ambiente dall'inquinamento atmosferico), delle quali è la pratica, e la stratificazione nel tempo delle evidenze scientifiche, convertite in normativa di dettaglio, a dettare i confini. Pare allora legittimo, forse, guardare con qualche sospetto a determinate critiche alle iniziative di regolazione dell'Unione Europea che sembrano sottointendere, nella sostanza, che l'assenza di regolazione sarebbe preferibile per ragioni di tutela del mercato[551]. Si tratterebbe, infatti, di un approccio che finirebbe per non affrontare problemi già attuali e concreti, di cui la discriminazione è solo un esempio, in grado di avere un impatto significativo sui diritti fondamentali dell'individuo, e, in ultima analisi, di limitare lo sviluppo tecnologico sul lungo termine, poiché le conseguenze che quest'ultimo potrebbe generare, in un contesto di totale *deregulation*, finirebbero per apparire inaccettabili nel contesto europeo, a livello culturale prima che giuridico. Né può sostenersi che lo sviluppo di una normativa di settore sia di per sé sempre un

[550] Per i rilievi da parte di tecnici e operatori del settore cfr. *supra*, n. 546. Per una critica d'ambito giuridico, v. invece D. LILKOV, *Regulating Artificial Intelligence in the EU: A Risky Game*, in *EuropeanView*, 20, 2, 2021, p. 166–174.

[551] Da questo punto di vista paiono significative, in particolare, le critiche alla Proposta di Regolamento mosse dal *think tank* CENTER FOR DATA INNOVATION durante la consultazione pubblica condotta dalla Commissione Europea nella seconda metà del 2021, consultabili all'indirizzo https://www2.datainnovation.org/2021-feedback-aia.pdf (8 agosto 2022), in cui si afferma, fin dalla prima pagina: «The AIA is too broad in its attempt to regulate an entire stack of technologies and applications at such an early stage in the development of AI. The added cost for the development and deployment of AI imposed by the many regulatory obligations in the Act will impose an expensive burden on the European digital ecosystem. In particular, the AIA, along with other regulatory barriers to market entry and growth, will make it difficult for European digital entrepreneurs to set up new businesses, grow them, and in the process create jobs, technological progress, and wealth».

limite a quello economico, posto che, in passato, tale assunto è stato più volte smentito dai fatti, come dimostra, ad esempio, la citata normativa a tutela dell'ambiente.

7. E quando l'intelligenza artificiale non discrimina? Una riflessione sulla diffusione della decisione automatizzata e le sue possibili conseguenze

Il lavoro più adatto alle proprie competenze, capacità e possibilità. Un reddito calcolato minuziosamente, corrispondente all'effettivo valore di tale lavoro. Una casa, un'auto e – perché no – un gruppo di amici o un partner sentimentale scelti con l'aiuto di algoritmi in grado di individuare, grazie all'analisi dei dati, "l'anima gemella" di ciascun essere umano. In poche parole, una vita in cui ogni scelta è guidata dal supporto di tecnologie che, per definizione, ci conoscono meglio di noi stessi. E in cui ogni decisione di terzi che ci riguardi è presa nello stesso modo.

In fondo, si tratta, almeno in parte, di un possibile scenario futuro, conseguenza dello sviluppo tecnologico. L'utilizzo di algoritmi a supporto o in sostituzione dei processi decisionali, infatti, è destinato a diffondersi sempre di più.Come visto, il dibattito sui rischi connessi a queste trasformazoni si è concentrato, finora, sulle conseguenze di possibili difetti o inefficienze nel loro funzionamento e utilizzo, in primo luogo discriminatorie. A conclusione di questo capitolo, pare opportuno svolgere alcune considerazioni di carattere metagiuridico su un tema diverso: i mutamenti causati dalle tecnologie in esame quando sono perfettamente funzionanti.

Supponiamo, ad esempio, che l'analisi dei dati metta in luce, con un inedito livello di precisione, differenze nell'attitudine a svolgere determinate mansioni o nella propensione a delinquere sulla base del genere, delle condizioni patrimoniali, dell'etnia. Differenze prima non note o considerate semplici pregiudizi. Oppure si immagini una società in cui gran parte delle diseguaglianze, in primo luogo di reddito, trovino una giustificazione - certamente meramente statistica, ma estremamente solida -nella valutazione di un algoritmo. Quali conseguenze potrebbero derivarne? "Misurare" algoritmicamente la diseguaglianza non conduce necessariamente a implementare misure per mettervi fine. Le ipotizzate differenze in base al genere nella propensione a delinquere verrebbero imputate a ragioni socio-culturali, con l'introduzione di apposite misure, in primo luogo in campo educativo, o poste a giustificazione di un trattamento differenziato, meno garantista verso una determinata categoria, da parte del sistema penale? Quali conseguenze potrebbe avere, dal punto di vista della solidarietà sociale tra i cittadini, lo sviluppo di una società in cui la diseguaglianza acquisti un connotato di oggettività tecnico-scientifica? Non può dimenticarsi che la costruzione di teorie che giustificassero razionalmente la discriminazione di determinate categorie, in primo luogo su base etnica, è stata tra le principali preoccupazioni di diversi regimi totalitari, e ha fornito una base di legittimazione culturale alle persecuzioni avvenute nel corso del XX secolo. La prospettiva

della società algoritmica, inoltre, suscita riflessioni anche adottando una prospettiva individualista: la diffusione di valutazioni algoritmiche sempre più raffinate come influirà sulla voglia del singolo di migliorarsi? Sarà facile come oggi ritenersi capaci di raggiungere determinati obiettivi, e impegnarsi nel farlo, di fronte a elaborazioni tecnologiche che sembrano indicare il contrario?

In breve, non va trascurato quanto l'errore, la tenacia, la solidarietà nel perdonare e correggere le reciproche imperfezioni abbiano un ruolo nella nostra società, influenzando il modo in cui sono distribuite le diseguglianze. Se ècerto che ciò rappresenta un limite a una piena realizzazione dell'eguaglianza di partenza, frustrata dalla difesa di interessi familari, amicali o corporativi, è anche vero che, alla base di tale forma di organizzazione sociale, vi è una componente culturale di reciproca comprensione. Si tratta di uno degli elementi che rendono più *umana* la nostra società, aggettivo che, non a caso, è spesso utilizzato per indicare un atteggiamento di benevolenza verso le incoerenze e le imperfezioni altrui. Una società in cui gli algoritmi abbiano un peso crescente nelle decisioni degli individui che la formano sarebbe di certo più efficiente, produttiva e, in moltissimi casi, più giusta. Non si può ignorare, però, il rischio che essa, talvolta, si riveli anche una società più crudele. E, come già detto, nonsolonel caso di eventuali difetti degli algoritmi, ma anche in ragione delle conseguenze derivanti dal loro corretto funzionamento. Un aspetto che a volte sembra trascurato dalla riflessione giuridica, ma anche filosofica e sociologica, e che pare invece cruciale per una comprensione precoce delle dinamiche che animeranno la "società algoritmica" che verrà[552].

[552] Sulle tematiche qui brevemente tratteggiate è di particolare interesse il contributo di P. BENANTI, *Oracoli. Tra algoretica e algocrazia*, Roma, 2018.Tra le riflessioni riguardanti, in senso ampio, le possibili ripercussioni a lungo termine dell'utilizzo di algoritmi nei processi decisionali, e in generale dello sviluppo tecnologico, non possono non segnalarsi le opere di Y. N. HARARI, in particolare *Homo deus: a brief history of tomorrow*, Gerusalemme 2015, e L. FLORIDI, in particolare *The fourth revolution*, Oxford, 2017.

PARTE III
Intelligenza artificiale e nuovi diritti fondamentali

Nuove sfide per nuovi problemi. L'avvento dell'intelligenza artificiale e l'emersione di nuovi diritti

1. Il concetto di diritto fondamentale e il dibattito teorico sui nuovi diritti

1.1. La definizione di "diritto" nel diritto costituzionale e nelle altre discipline giuridiche

I diritti hanno, ovviamente, un ruolo centrale nella scienza del diritto costituzionale[553]. La base del costituzionalismo contemporaneo, l'art. 16 della Dichiarazione dei diritti dell'uomo e del cittadino, identifica nella tutela dei diritti e nella separazione dei poteri gli elementi essenziali perché uno stato possa dire di *avere una costituzione*[554]. Lo studio dei diritti può definirsi, allora, la "parte sostanziale" del diritto costituzionale, a cui si affianca una parte organizzativa e procedurale, che, attraverso lo studio delle forme di stato e governo, verifica l'effettività del principio della separazione dei poteri[555].

Allo stesso tempo, deve evidenziarsi che l'interesse per i diritti non è di certo un'esclusiva del diritto costituzionale. Non vi è, in realtà, disciplina giuridica che non abbia i diritti della persona come oggetto di studio: dal diritto penale (che incrimina le lesioni di beni giuridici cui corrispondono, in molti casi, diritti primari) al diritto processuale (le cui norme non sono che la

[553] Tra gli innumerevoli studi sui diritti possono indicarsi, da vari punti di vista e senz'animo di completezza, P. VIRGA, *Libertà giuridica e diritti fondamentali*, Milano, 1947; C. LAVAGNA, *Basi per uno studio delle figure giuridiche soggettive contenute nella Costituzione italiana*, Padova, 1953; E. CASETTA, *Diritti pubblici subiettivi*, in *Enciclopedia del diritto*, Milano, 1964, XII 791 ss.;R. DWORKIN, *Taking rights seriously*, Cambridge (US), 1977; P. BARILE, *Diritti dell'uomo e libertà fondamentali*, Bologna, 1984; G. JELLINEK, *La dichiarazione dei diritti dell'uomo e del cittadino* (1985), Bari, 2002;R. ALEXY, *Teoria dei diritti fondamentali* (1994), Bologna, 2012;G. PINO, *Diritti e interpretazione*, Bologna, 2010 e *Il costituzionalismo dei diritti*, Bologna, 2017;N. BOBBIO, *L'età dei diritti*, Torino, 1990; G. PECES-BARBA MARTÍNEZ, *Curso de derechos fondamentales. Teoría general*, Madrid, 1991; P. F. GROSSI, *Introduzione allo studio dei diritti inviolabili nella Costituzione italiana*, Padova, 1972; A. BALDASSARRE, *Diritti della persona e valori costituzionali*, Torino, 1997;L. FERRAIOLI, *Diritti fondamentali. Un dibattito teorico*, Bari, 2001 e *La democrazia attraverso i diritti*, Bari, 2013; A. PACE, *Problematica delle libertà costituzionali*, Padova, 2003; G. ROLLA, *La tutela dei diritti fondamentali*, Roma, 2012; F. CORTESE, D. BORGONOVO RE, D. FLORENZANO, *Diritti inviolabili, doveri di solidarietà e principio di eguaglianza*, 2015; P. CARETTI, G. TARLI BARBERI, *I diritti fondamentali*, Torino, 2017; R. BIN, *Critica della teoria dei diritti*, Milano, 2018.

[554] L'art. 16 della *Declaration* del 1789, com'è noto, recita: «Toute société dans laquelle la garantie des droits n'est pas assurée, ni la séparation des pouvoirs déterminée, n'a point de constitution». Tra i moltissimi commenti, cfr., ad esempio, P. ALBERTINI, *Article 16*, in G. CONAC, M. DEBENE, G. TEBOUL, *La déclaration des droits de l'homme et du citoyen de 1789, Histoire analyse et commentaires*, Parigi, 1993, p. 331 ss.;M. BOUAZIZ, *Significations et interprétations de l'article 16 de la Déclaration des droits de l'homme et du citoyen de 1789. Contribution à l'histoire de la notion de constitution,* Parigi, 2019.

[555] Sistematizza in questo modo la disciplina, ad esempio, M. OLIVETTI, *Diritti fondamentali*, Torino, 2020, p. 3.

procedimentalizzazione delle garanzie dell'individuo), dal diritto civile e commerciale (per i diritti patrimoniali) al diritto del lavoro (in cui si sostanziano numerosi dei c.d. diritti sociali).In questo scenario, può non essere agevole distinguere quali siano i diritti verso i quali il diritto costituzionale concentra il proprio interesse, e cosa li caratterizzi.

La nozione generica, interdisciplinare di diritti li intende come situazioni giuridiche soggettive attive, di vantaggio, tutelate dall'ordinamento [556]. Si tratta di una definizione in grado di ricomprendere, in potenza, ogni pretesa individuale nei confronti dei pubblici poteri riconosciuta dalla legge e i diritti patrimoniali tra privati, sia relativi che assoluti. Relativamente a cosa distingua, invece, i diritti *fondamentali*, che interessano il diritto costituzionale, le innumerevoli elaborazioni teoriche possono riassumersi, in via di estrema semplificazione, in due filoni[557]. Secondo una prima corrente di pensiero, i diritti sono fondamentali per la fonte che li individua (la Costituzione, legge fondamentale dello stato)[558]; secondo un'altra ricostruzione, a identificare i diritti fondamentali è in primo luogo il contenuto, a prescindere dalla loro positivizzazione[559]. Quest'ultima tesi lascia aperta la discussione sulle modalità di determinazione di tale contenuto dei diritti fondamentali: accanto a concezioni neogiusnaturalistiche, vi sono visioni più moderne, che individuano come decisivi elementi strutturali, ad esempio la potenziale *universalità* di un diritto, intesa come possibilità di essere riconosciuto, in astratto, in ugual misura ad ogni individuo, a differenza dei diritti patrimoniali (è una tesi portata avanti, nella dottrina italiana, in alcuni lavori di Luigi Ferrajoli)[560].

Si tratta, in ogni caso, di distinzioni che non vanno sopravvalutate, eredità ultima del dibattito tra giusnaturalismo e giuspositivismo sulla natura dei diritti, tanto risalente quanto inesauribile[561]. Non è oggi messo in discussione, infatti, che la definizione di diritto non possa fondarsi sul solo elemento formale della positivizzazione, e che il contenuto, anzi, abbia un ruolo centrale: i diritti fondamentali corrispondono sempre a garanzie ultime di protezione dell'individuo verso un

[556] Propone questa definizione ad esempio G. PINO, *Diritti e interpretazione*, Bologna, 2010, p. 12.

[557] L'espressione "diritti fondamentali" ha origine nella cultura giuridica tedesca, ed è impiegata nelle costituzioni di Francoforte (1848), Weimar (1919) e Bonn (1949), cfr. F. LANCHESTER, *Le costituzioni tedesche da Francoforte a Bonn. Introduzione e testi*, Milano, 2009. Com'è noto, la Costituzione italiana la utilizza una sola volta, all'art. 32, come attributo del diritto alla salute. Essa, invece, è relativamente frequente nella giurisprudenza della Corte costituzionale, specialmente negli ultimi decenni (si vedano, ad esempio, Corte cost. sent. n. 120/2014, in cui i diritti fondamentali sono individuati tra i possibili limiti all'autodichia delle Camere, o Corte cost. sent n. 63/2016, che definisce fondamentale il diritto alla libertà religiosa) e in dottrina e manualistica.

[558]Per un'affermazione netta in tal senso nella dottrina contemporanea, cfr.M. OLIVETTI, *Diritti fondamentali cit.*, p. 5.

[559]Tra gli autori più influenti degli ultimi decenni a vario titolo riconducibili a questa impostazione si rimanda a G. PECES BARBA, *Curso de derechos fondamentales. Teoría general*, p. 20 ss. e, come affermato nel testo e specificato alla nota successiva, Luigi Ferrajoli.

[560]Cfr. L. FERRAJOLI, *Diritti fondamentali*, *I diritti fondamentali nella teoria del diritto*, e *I fondamenti dei diritti fondamentali*, tutti raccolti in L. FERRAJOLI (A CURA DI), *Diritti fondamentali. Un dibattito teorico cit.*, rispettivamente p. 5 ss., 121 ss. e 279 ss. Il primo dei tre contributi è aperto da questa definizione: «sono "diritti fondamentali" tutti quei diritti soggettivi che spettano universalmente a tutti gli esseri umani in quanto dotati dello *status* di persone, o di cittadini o di persone capaci d'agire».

[561]*Ex multis* si rimanda, per l'autorevolezza, alle ricostruzioni di N. BOBBIO, *Sul fondamento dei diritti dell'uomo*, in *L'età dei diritti cit.*, p. 5 ss. e G. PECES BARBA, *Curso de derechos fondamentales. Teoría general*, p. 1 ss.

potere.Il ruolo delle positivizzazione rimane, comunque, essenziale, poiché ogni definizione di diritti fondamentali che guardi al loro contenuto sottende una scelta di valori che non è sottratta a discussione, e può trovare un forte argomento di legittimazione nel riconoscimento, almeno parziale, da parte della legge scritta. Il discorso sui diritti contemporaneo, dunque, è prima di tutto un discorso sulle Carte internazionali dei diritti che si sono susseguite a partire dalla rivoluzione francese e sui *bills of rights* codificati nelle diverse costituzioni[562]. Ne deriva che, se il discorso sui diritti era intrinsecamente interdisciplinare, lo stesso deve dirsi di quello sui diritti fondamentali, che non interessano solamente il diritto costituzionale. Anche se le due categorie non sono sovrapponibili, infatti, l'oggetto di studi della branca del diritto internazionale che si occupa di protezione dei diritti dell'uomo presenta, di certo, pesanti interferenze coi diritti fondamentali. Tanto che l'interazione tra costituzioni, carte internazionali e i relativi strumenti di protezione dei diritti ha dato luogo al fenomeno della c.d. tutela multivello, da tempo oggetto di interesse da parte degli studiosi delle due discipline[563]. Anche il tema dei *nuovi diritti*, cui questa terza parte del lavoro sarà dedicata, è stato al centro della riflessione sia internazionalistica che costituzionalistica.

[562] In epoca contemporanea, il riferimento, sul piano internazionale, è, in primo luogo, alla Dichiarazione universale dei diritti umani delle Nazioni Unite del 10 dicembre 1948 e alle successive Convenzioni sui diritti economici, sociali e culturali e sui diritti civili e politici del 1966. Non possono non menzionarsi, inoltre, la Convenzione Europea dei Diritti dell'Uomo del 1950 e le altre Carte regionali dei diritti che vi hanno fatto seguito, come la Convenzione interamericana del 1969, la Carta africana del 1981 e la Carta araba del 2004, le quali, com'è stato sottolineato da più parti, mettono in luce le forti differenze nella concezione dei diritti che possono derivare da tradizioni politico-culturali distinte. Per alcuni commenti cfr., da varie prospettive, S. BARTOLE, B. CONFORTI, G. RAIMONDI, *Commentario alla Convenzione europea dei diritti dell'uomo e delle libertà fondamentali*, Padova, 2001; S. BARTOLE, P. DE SENA, V. ZAGREBELSKY (A CURA DI), *Commentario breve alla Convenzione europea dei diritti dell'uomo*, Padova, 2012; M. CURCIO, *La dichiarazione dei diritti delle Nazioni Unite*, Milano, 1950; L. D'ANDREA, G. MOSCHELLA, A. RUGGERI, A. SAITTA (A CURA DI), *La Carta dei diritti dell'Unione europea e le altre Carte (ascendenze culturali e mutue implicazioni)*, Torino, 2016; R. MURRAY, *The African Charter of Human and Peoples' Rights. A commentary*, Oxford, 2019; A. M. EL-DEMERY, *The arab charter of human rights: a voice for sharia in the modern world*, Indianapolis, 2015. Per il riconoscimento dei diritti nei diversi sistemi costituzionali – e la loro retrocessione, in vari momenti della storia contemporanea – si rimanda, in primo luogo, all'analisi globale condotta da Samuel P. Huntington nel noto S. P. HUNTINGTON, *The third wawe: democratization in the late twentieth century*, Norman, 1991.

[563] Nella vastissima letteratura giuridica italiana sul tema possono richiamarsi, senz'animo di completezza e da punti di vista anche molto diversi, A. CARDONE, *La tutela multilivello dei diritti fondamentali*, Milano, 2012; L. PINESCHI, *Diritti umani (protezione internazionale dei)*, in *Enciclopedia del diritto – Annali*, V, 2012; A. CASSESE, *I diritti umani oggi*, Roma, 2008; M. CARTABIA (A CURA DI), *I diritti in azione*, Bologna, 2007; A. MARCHESI, *La protezione internazionale dei diritti umani*, Torino, 2021; S. ZAPPALÀ, *La tutela internazionale dei diritti umani*, Bologna, 2011; L. MEZZETTI, *Diritti umani: protezione internazionale e ordinamenti nazionali*, Pisa, 2021; F. G. PIZZETTI, *Aspetti e problemi del costituzionalismo multilivello*, Milano, 2004; A. D'ANTENA, *Tutela dei diritti fondamentali e costituzionalismo multilivello: tra Europa e stati nazionali*, Milano, 2004; R. BIFULCO, M. CARTABIA, A. CELOTTO (A CURA DI), *L'Europa dei diritti*, Bologna, 2001; BILANCIA P., DE MARCO E., PIZZETTI F.G., *"Nuovi diritti" e "tutela multilivello dei diritti",* in BILANCIA P., DE MARCO E., PIZZETTI F.G. (a cura di), *L'ordinamento della Repubblica: le istituzioni e la società*, Padova, 2021, p. 517 ss.*;* D. BUTTURINI, *La tutela dei diritti fondamentali nell'ordinamento costituzionale italiano ed europeo*, Napoli, 2009; L. MONTANARI, *I diritti dell'uomo nell'area europea tra fonti internazionali e fonti interne*, Torino, 2002; B. RANDAZZO, *Giustizia costituzionale sovranazionale. La Corte europea dei diritti dell'uomo*, Milano, 2012; L. CASSETTI (A CURA DI), *Diritti, principi e garanzie sotto la lente dei giudici di Strasburgo*, Napoli, 2012; D. TEGA, *I diritti in crisi. Tra Corti nazionali e Corte europea di Strasburgo*, Milano, 2012. Tra le moltissime opere di studiosi stranieri, invece, si indicano E. U. PETERSMANN, *Multilevel constitutionalism for multilevel governance of public goods*, Oxford-Portland, 2017; N. CHRONOWSKI, *Human rights in a multilevel constitutional area*, Parigi, 2018; F. G. ISA; K. DE FEYTER, *International protection of human rights: achievements and*

1.2 I nuovi diritti nel diritto internazionale e nel diritto costituzionale

Il diritto internazionale non ha mai concepito i diritti umani come un sistema chiuso, refrattario ad aggiornamenti. La dottrina internazionalistica ha a lungo discusso, e discute ancora oggi, di "generazioni" di diritti umani, prendendo spunto da un celebre intervento di Karel Vasak all'Istituto internazionale per i diritti dell'uomo di Strasburgo, nel 1979[564]. Nel tentativo di storicizzare e sistematizzare l'origine dei diversi diritti umani, Vasak identificava: una prima generazione di diritti, consistenti nei diritti civili e politici, la cui tutela impone ai poteri pubblici prevalentemente obblighi di astensione; una seconda generazione, formata dai diritti economici, sociali e culturali, la cui protezione implica, per lo stato, prestazioni positive; una terza generazione, composta dai c.d. diritti di solidarietà, la cui titolarità è spesso superindividuale e condivisa con le generazioni future: è il caso, tra gli altri, dei diritti all'ambiente, alla conservazione del patrimonio culturale, allo sviluppo, all'equità intergenerazionale. A queste potrebbe aggiungersi, a detta di diversi studiosi, una quarta generazione, composta da diritti a protezione di nuove istanze sociali strettamente connesse all'innovazione tecnologica. Si tratterebbe, ad esempio, della tutela dell'autodeterminazione nelle fasi terminali dell'esistenza, del diritto all'integrità del genoma umano di fronte a tecniche di ingegneria genetica, della protezione dalle forme più avanzate di sorveglianza[565].

Le generazioni di Vasak, mutate le circostanze, sono parte anche del patrimonio concettuale del diritto costituzionale. I diritti di prima generazione, infatti, corrispondono in larga parte alle libertà negative tutelate dallo stato liberale, e i diritti di seconda generazione allo stato sociale sviluppato,

---

challenges, Bilbao, 2006; R. ERGEC, M. HAPPOLD, *Protection européenne et internationale des droits de l'homme*, Bruxelles, 2014.

[564]K. VASAK, *A 30 years struggle. The sustained efforts to give force of law to the Universal Declaration of Human Rights*, The UNESCO Courier, Nov. 1977, p. 29-30. Per un commento sull'influenza della categorizzazione di Vasak sulla dottrina internazionalistica contemporanea, cfr. S. DOMARADZKI, M. KHVOSTOVA, D. PUPOVAC, *Karel Vasak's Generations of Rights and the Contemporary Human Rights Discourse*, in *Human Rights Review*, 20, 2019, p. 423-443.

[565] Identificano i diritti di quarta generazione con le istanze di protezione generate dallo sviluppo tecnologico e biomedico, ad esempio, E. CECCHERINI, *La codificazione dei diritti nelle recenti costituzioni*, Milano, 2002, p. 122 ss.; J. C. RIOFRÍO MARTÍNEZ VILLALBA, *La cuarta ola de derechos humanos: los derechos digitales*, in *Revista latinoamericana de derechos humanos*, 25, 1, 2014, p. 15-45; J. BUSTAMANTE, *Hacia la cuarta generación de Derechos Humanos: repensando la condición humana en la sociedad tecnologica*, in *Revista iberoamericana de ciencia, tecnología, sociedad e innovación*, 1, 2001; E. VALDES, *Biolaw, genetic harm and fourth generation human rights*, in *Boletín mexicano de derecho comparado*, 48, 144, 2015, p. 1197-1228. Altri autori, invece, hanno posto l'accento sui diritti delle generazioni future e su valori come la democrazia, l'informazione e il pluralismo, che acquisirebbero il rango di veri e propri diritti con l'avvento della quarta generazione, cfr. rispettivamente R. BIFULCO, *Diritto e generazioni future. Problemi giuridici della responsabilità intergenerazionale*, Milano, 2008 e P. BONAVIDES, *Curso de direito constitucional,* XXVIII ed., Malheiros (Brasile), 2013, p. 590 ss. Non può non evidenziarsi, infine, che alcune costituzioni frutto delle transizioni democratiche più recenti riconoscono lo status di diritti fondamentali a beni ancora ulteriori, come casa, cibo, o acqua, che potrebbero anch'essi ricondursi, sul piano teorico, a un'eventuale quarta generazione di diritti. Si tratta, ad esempio, dell'art. 16 della Costituzione brasiliana, riformato nel 2010 per includervi un riferimento all'alimentazione, e dell'art. 44 della Costituzione tunisina del 2014, per quanto riguarda l'accesso all'acqua.

nel secondo dopoguerra, negli stati costituzionali di diritto. Ferma questa distinzione, deve rilevarsi, però, che la possibilità di teorizzare diritti ulteriori è stata sottoposta dalla dottrina costituzionalistica a un vaglio critico molto più profondo di quanto fatto da quella internazionalistica. In particolare, la letteratura giuridica italiana si è a lungo divisa tra chi considerava il catalogo dei diritti esplicitato dalla Carta un elenco chiuso e chi vedeva nell'art. 2 della Costituzione un'apertura al riconoscimento di diritti fondamentali ulteriori rispetto a quelli codificati dai costituenti[566]. Dopo decenni di diffidenza, la tesi della "fattispecie aperta" è oggi prevalente e ha ricevuto riconoscimento anche da parte della Corte costituzionale, almeno a partire dalla nota sentenza n. 561 del 1987[567]. In tale pronuncia, la Consulta ha individuato direttamente nell'art. 2 la fonte della protezione del diritto all'integrità sessuale, non esplicitato nella Carta, ricavandone l'illegittimità costituzionale del mancato risarcimento, per mezzo di specifiche pensioni di guerra, del danno morale derivante dalle violenze carnali compiute dalle truppe straniere impegnate nella campagna d'Italia nel corso della seconda guerra mondiale. Da allora, il Giudice delle Leggi ha comunque fatto un uso prudente dell'"apertura" dell'art. 2: valga, come esempio, l'altrettanto nota sentenza n. 138 del 2010, in cui, pur includendo le unioni omosessuali tra le formazioni sociali tutelate dalla norma, la Corte ha escluso che da ciò potesse derivare, senza un intervento del Legislatore, il diritto a contrarre matrimonio[568]. In ogni caso, deve tenersi presente

[566] Tra i sostenitori della tesi della "fattispecie aperta" possono ricordarsi, ad esempio, C. MORTATI, *La Corte costituzionale e i presupposti della sua vitalità*, in *Iustitia*, 1949, p. 69 ss., con accezione neogiusnaturalistica, e A. BARBERA, *Art. 2*, in G. BRANCA (A CURA DI), *Commentario della Costituzione*, Bologna-Roma, 1975, p. 50 ss., con un'interpretazione che valorizza l'evoluzione dei costumi. Prendono posizione per la fattispecie chiusa, interpretando l'art. 2 come un rinvio ai diritti che la Costituzione riconosce e disciplina, ad esempio, M. MAZIOTTI DI CELSO, *Lezioni di diritto costituzionale*, II, Milano, 1985, p. 57 ss. e M. OLIVETTI, *Diritti fondamentali cit.*, p. 149-151. Evidenziano possibili incoerenze della teoria dell'apertura anche P. BARILE, *Diritti dell'uomo e libertà fondamentali cit.*, p. 52, che offre la nota ricostruzione dell'art. 2 come *matrice* dell'inviolabilità dei diritti enumerati in Costituzione e non come *fonte* di nuovi diritti, e A. PACE, *Problematica delle libertà costituzionali. Parte generale*, 2003,p. 21 ss. Sul tema si vedano, da vari punti di vista, P. CARETTI, *I diritti fondamentali. Libertà e Diritti sociali*, Torino, 2005, p. 137 ss.; G. LOMBARDI, *Contributo allo studio dei doveri costituzionali*, Milano, 1967, p. 34 ss.; F. MODUGNO, *I nuovi diritti nella giurisprudenza costituzionale*, Torino, 1995; M. CARTABIA, *I "nuovi" diritti*, in *Stato, chiese e pluralismo confessionale*, 2, 2011; R. BIN, *Critica della teoria dei diritti*, Milano, 2018, p. 55 ss.

[567] In particolare, nella sentenza si legge, al punto n. 2 del *Considerato in diritto*: «Essendo la sessualità uno degli essenziali modi di espressione della persona umana, il diritto di disporne liberamente è senza dubbio un diritto soggettivo assoluto, che va ricompreso tra le posizioni soggettive direttamente tutelate dalla Costituzione ed inquadrato tra i diritti inviolabili della persona umana che l'art. 2 Cost. impone di garantire».

[568] Propone questa lettura M. CARTABIA, *I "nuovi" diritti cit.*, p. 2-3. Nello specifico, la sentenza 138/2010, pur riconoscendo che l'art. 2 Cost. garantisce copertura costituzionale anche al diritto fondamentale a vivere liberamente la condizione di coppia omossessuale, da annoverarsi tra le formazioni sociali, ha ritenuto che le modalità concrete di tutela di tale diritto, tra le quali l'estensione della disciplina del matrimonio è solo una delle possibili ipotesi, rientrino nella discrezionalità del Legislatore: «L'art. 2 Cost. dispone che la Repubblica riconosce e garantisce i diritti inviolabili dell'uomo, sia come singolo sia nelle formazioni sociali ove si svolge la sua personalità e richiede l'adempimento dei doveri inderogabili di solidarietà politica, economica e sociale. Orbene, per formazione sociale deve intendersi ogni forma di comunità, semplice o complessa, idonea a consentire e favorire il libero sviluppo della persona nella vita di relazione, nel contesto di una valorizzazione del modello pluralistico. In tale nozione è da annoverare anche l'unione omosessuale, intesa come stabile convivenza tra due persone dello stesso sesso, cui spetta il diritto fondamentale di vivere liberamente una condizione di coppia, ottenendone – nei tempi, nei modi e nei limiti stabiliti dalla legge – il riconoscimento giuridico con i connessi diritti e doveri. Si deve escludere, tuttavia, che l'aspirazione a tale riconoscimento – che necessariamente postula una disciplina di carattere generale, finalizzata a regolare diritti e doveri

che il dibattito tra teorici dell'art. 2 come clausola aperta o chiusa ha portata pratica relativa:generalmente, infatti, i sostenitori della "chiusura"riconoscono comunque la possibilità di dare copertura costituzionale a istanze innovative, frutto dell'evoluzione sociale, riconducendole all'ambito applicativo di diritti esplicitati nella Carta, interpretati estensivamente[569].

Volgendo lo sguardo alla menzionata tutela multilivello, la Corte EDU si è dimostrata molto cauta nel riconoscere nuovi diritti, vedendovi il rischio di invadere il margine di apprezzamento degli stati[570]. In ogni caso, essa è stata protagonista, nel corso degli anni, di aperture significative, specialmente attraverso un'interpretazione estensiva del diritto alla vita privata e familiare previsto all'art. 8 della Convenzione[571]. La possibilità di riconoscere situazioni giuridiche frutto dell'evoluzione dei costumi, della società e delle tecnologie, inoltre, si è presentata nella riflessione costituzionale anche di altri paesi democratici; diverse costituzioni, peraltro, risolvono in radice la questione, incorporando clausole di apertura a diritti non codificati molto più esplicite di quella italiana (è il caso, ad esempio, del IX Emendamento della Costituzione degli Stati Uniti)[572].

---

dei componenti della coppia – possa essere realizzata soltanto attraverso una equiparazione delle unioni omosessuali al matrimonio. È sufficiente l'esame, anche non esaustivo, delle legislazioni dei Paesi che finora hanno riconosciuto le unioni suddette per verificare la diversità delle scelte operate», *Considerato in diritto*, punto n. 8.

[569] Sul punto si vedano le considerazioni di R. BIN, *Nuovi diritti e vecchie questioni*, in A.A.V.V., *Studi in onore di Luigi Costato*, III, Napoli, 2014, p. 75-84 e R. BIN, *Critica della teoria dei diritti cit.*, p. 55 ss, che invita a riflettere su come, in molti casi, le situazioni giuridiche dibattute in termini di "nuovi diritti" siano inquadrabili, più semplicemente, come nuove forme di manifestazione di "vecchie" libertà. Quanto affermato in materia di rilievo principalmente teorico del dibattito tra fattispecie aperta e fattispecie chiusa è vero, comunque, solo con importanti specificazioni: deve segnalarsi, in particolare, la posizione dei citati M. MAZIOTTI DI CELSO, *Lezioni di diritto costituzionale cit.*, 1985, p. 57-58 e M. OLIVETTI, *Diritti fondamentali cit.*, p. 149-151, che, senza mettere in discussione la possibilità di interpretare le norme della Costituzione in base alle regole interpretative ordinarie (e dunque, per ipotesi, anche estensivamente), considerano, però, uno dei motivi essenziali per rigettare la tesi della fattispecie aperta proprio l'allargamento delle prerogative costituzionali attraverso l'interpretazione, in primo luogo giudiziale, cui essa porta e che è, invece, preferibile evitare. Secondo i due autori, la teoria dell'apertura, combinata con la tesi che include i diritti inviolabili tra i limiti impliciti della revisione costituzionale, condurrebbe al non auspicabile risultato di sottrarre gli eventuali nuovi diritti non solo all'azione riformatrice del Legislatore ordinario, ma anche alla stessa revisione costituzionale.

[570] Paradigmatica, da questo punto di vista, la vicenda del diritto alla genitorialità e della compatibilità con la Convenzione EDU del divieto di procreazione medicalmente assistita per mezzo di fecondazione eterologa, considerato dalla Prima Sezione della Corte una violazione dell'art. 8 della Carta nella sentenza *S. H. et al. v. Austria* (57813/00), 1 aprile 2010, con un'interpretazione poi smentita dalla Grande Chambre in *S. H. et al. v. Austria* (57813/00), 3 novembre 2011, che ha sancito che la disciplina della fecondazione eterologa rientra nel margine di apprezzamento degli stati firmatari.

[571] Ciò ha portato al riconoscimento di prerogative sorte con l'evoluzione dei costumi e di certo non considerate al momento della stipulazione della Convenzione. La Grande Chambre della Corte, ad esempio, ha statuito che l'art. 8 della Convenzione impone di garantire adeguate forme di riconoscimento legale delle coppie omoaffettive, nonostante l'art. 12 attribuisca il diritto di contrarre matrimonio solamente a "uomini e donne", cfr. casi *Vallianatos et al. v. Greece* (7 novembre 2013) e *Oliari et al. v. Italia* (18766/11 e 36030/11), 21 luglio 2015. Sempre in base all'art. 8, in combinato disposto col divieto di discriminazione di cui all'art. 14, la Grande Chambre ha sancito l'incompatibilità con la Convenzione di discriminazioni tra coppie non sposate sulla sola base dell'orientamento sessuale in materia di adozione del figlio del partner, in *X et al. v. Austria* (19010/07), 19 febbraio 2013, e di differenziazioni tra persone omosessuali ed eterosessuali nei casi in cui un ordinamento consenta l'adozione al singolo individuo, in *E. B. v. France* (43546/02), 22 gennaio 2008.

[572] Com'è noto, il Nono Emendamento recita: «The enumeration in the Constitution of certain rights shall not be construed to deny or disparage others retained by the people». Negli anni '60, '70 e '80 del Novecento, la Corte Suprema ha più volte utilizzato il IX Emendamento nelle sue argomentazioni, al fine di riconoscere protezione giuridica a diritti individuali non esplicitati nella Costituzione americana, spesso congiuntamente al *substantive due process* protetto dal V e XIV Emendamento. Sono basate su questi emendamenti, ad esempio, le note pronunce *Griswold v.*

1.3 Come nasce un nuovo diritto? Un'ipotesi teorica su intelligenza artificiale e nuovi diritti

Come visto al paragrafo precedente, il discorso sull'ammissibilità di nuovi diritti può dirsi ormai risolto in termini positivi. Questione diversa, e ben più controversa, è invece definire quali caratteristiche rendano una determinata situazione giuridica di recente apparizione un nuovo diritto fondamentale. Nell'ottica scelta da questo lavoro, è prima di tutto il contenuto a venire in rilievo.

Un *nuovo diritto* è una nuova istanza sociale, riconoscibile in egual misura in capo ad ogni persona e generata dal mutamento dei costumi o dall'innovazione tecnologica, relativa alla protezione dell'individuo nei confronti di un potere, talvolta nuovo anch'esso.L'eventuale positivizzazione non può essere un elemento decisivo: l'attesa dell'intervento del Legislatore, magari a livello costituzionale, o dal punto di vista della tutela internazionale dei diritti umani, porterebbe a prolungati vuoti di tutela proprio nell'ambito in cui essa è vitale, i diritti fondamentali, e di fronte alle istanze percepite più urgenti, perché generate dal cambiamento e relative a settori in cui il diritto scritto si dimostra particolarmente deficitario. Non può trascurarsi, inoltre, che un *nuovo diritto*, una volta positivizzato, non è più tale, perché entra nel catalogo di quelli pacificamente ed esplicitamente riconosciuti: ogni teoria che accetti la possibilità che nuovi diritti emergano col mutare di tempi e contesti, quindi, non può considerare la positivizzazione in costituzioni e Carte dei diritti un elemento necessario per la loro configurazione.

Allo stesso tempo, il diritto scritto ha di certo un ruolo importante nel riconoscimento di protezione a nuove situazioni giuridiche soggettive. In molti casi, un nuovo diritto nasce nella coscienza sociale, per poi trovare riconoscimento e protezione prima nelle sentenze dei giudici, attraverso l'interpretazione estensiva, analogica o apertamente creatrice,poi in norme di fonte primaria o secondaria. Solo raramente, e a distanza di molto tempo, trova ingresso esplicito in carte e costituzioni[573]. Il diritto di rango sub-costituzionale, quindi, può fungere da utile indice per

Connecticut (1965) in materia di contraccettivi e *Roe v. Wade* (1973) in materia di aborto, di recente oggetto di *overruling* con la pronuncia *Dobbs v. Jackson Women's Health Organization*(2022), che, contenendo affermazioni molto nette sulla necessità di riconoscere protezione a diritti ulteriori rispetto a quelli inclusi nel *Bill of Rights* solo qualora siano fortemente radicati nella "storia della Nazione", sembra rappresentare un radicale cambio di rotta riguardo alla concezione aperta dei citati Emendamenti V, IX e XIV. Per dei commenti v. R. BERGER, *Ninth Amendment*, in *Cornell Law Review*, 1, 1980, p. 1-26; R. L. CAPLAN, *The History and Meaning of the Ninth Amendment*, in *Virginia Law Review*, 69, 2, 1983, p. 223-268.

[573] Interessante, da questo punto di vista, è la vicenda del bene ambiente – spesso citato, in ambito internazionalistico e costituzionale, tra i nuovi diritti di natura superindividuale - nell'ordinamento italiano. Forme di protezione del patrimonio ecologico e ambientale, con norme tecniche e sanzioni amministrative o penali, sono, infatti, molto risalenti, e il diritto dell'ambiente è da decenni considerato una corposa disciplina giuridica a sé stante. A testimonianza della complessità della materia, la sistematizzazione in un Testo Unico delle norme sull'argomento, ad opera del c.d. *Codice dell'ambiente*, risale al d.lgs. n. 152 del 3 aprile 2006. L'esplicitazione in costituzione della sua protezione, però, è avvenuta solo con la recente Legge costituzionale n. 1 dell'11 febbraio 2022, dopo numerosi anni in cui la dottrina giuridica ne sottolineava il rilievo al massimo livello. Per dei commenti in letteratura, cfr. L. BARTOLUCCI, *Le generazioni future (con la tutela dell'ambiente) entrano "espressamente" in Costituzione*, in *Forum di Quaderni Costituzionali*, 2, 2022, p. 20 ss.; I. RIVERA, *Le tonalità dell'ambiente e le generazioni future nel cammino di riforma della Costituzione*, in *BioLaw Journal - Rivista di BioDiritto*, 2, 2022, p. 225 ss. e, in chiave critica, D. AMIRANTE, *La reformette dell'ambiente in Italia e le ambizioni del costituzionalismo ambientale*, in *Diritto pubblico comparato ed*

l'identificazione di una nuova situazione giuridica meritevole di protezione, che, per il rango primario dei beni coinvolti e le caratteristiche dell'istanza sociale da cui deriva, si eleva allo status di vero e proprio diritto fondamentale.

Quest'ultimo argomento avrà un ruolo centrale nel seguito di questo lavoro, dedicato all'analisi dei possibili nuovi diritti fondamentali connessi all'era dell'intelligenza artificiale. I prossimi capitoli, infatti, saranno dedicati all'esame di alcune nascenti istanze sociali, generate dall'evoluzione tecnologica, e in particolare dall'avvento dell'intelligenza artificiale, che, per il ruolo di garanzia individuale dalle possibili conseguenze di determinate applicazioni e utilizzi di tali tecnologie, paiono destinate a configurarsi come veri e propri nuovi diritti. Un indice del loro crescente riconoscimento sarà individuato anche nelle prime ipotesi di regolazione contenenti presidi a loro tutela, che saranno analizzate in profondità. Si tratta, per l'appunto, di testi normativi di rango primario o secondario, spesso di natura tecnica, a cominciare dalla Proposta di Regolamento in materia di intelligenza artificiale della Commissione Europea. I possibili nuovi diritti cui si fa riferimento sono, come si dirà: il diritto a una sorta di *AI disclosure*, ossia a conoscere la natura, umana o artificiale, di un interlocutore; il diritto a una spiegazione dei risultati di un sistema; il diritto a una soglia minima di controllo umano sulla tecnologia[574].

---

europeo, 2, 2022, p. 5 ss. Inoltre, la funzione delle norme di rango primario e subprimario per il riconoscimento di nuovi diritti è sottolineata da L. FERRAJOLI, *Diritti fondamentali cit.*, p. 9 ss., che individua nelle garanzie disposte dalla legge la componente fondamentale dell'effettività dei diritti, anche se specifica che da tali garanzie essi vanno sempre tenuti distinti, e non è il riconoscimento nel diritto positivo di un ordinamento a determinarne l'esistenza. Sul tema cfr. anche, pur con un'impostazione molto diversa da quella qui portata avanti, M. OLIVETTI, *Diritti fondamentali cit.* p. 148-149, che, nella sua critica della visione dell'art. 2 come "fattispecie aperta", sottolinea indirettamente l'importanza della legge ordinaria per la tutela di nuove istanze fondamentali, evidenziando come debba essere il Parlamento, elettivo e rappresentativo, a riconoscere i nuovi diritti emersi nella coscienza sociale e non i giudici, ordinari o costituzionali, attraverso l'interpretazione dell'art. 2.

[574] Individua questi nuovi diritti C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Diritto pubblico comparato ed europeo*, numero speciale, 2019, p. 125 ss.; *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, in *BioLaw Journal – Rivista di BioDiritto*, Special Issue 2, 2019, p. 722 ss.; svolge considerazioni simili, pur esprimendosi in termini di principi e non di diritti, A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019, p. 63 ss. Di particolare interesse è anche la ricostruzione di L. M. AZZENA, *L'algoritmo nella formazione della decisione amministrativa: l'esperienza italiana*, in *Revista Brasileira de Estudos Políticos*, 123, 2021, p. 503-537, che, con un'impostazione in diversi punti accostabile a quella qui adottata, evidenzia che «i nuovi strumenti [le tecnologie di IA, ndr] possono risultare propulsori per l'affermazione di nuovi diritti e utili a rafforzare l'effettività di quelli già dati» e rileva come si stia formando «un nuovo catalogo di diritti […] secondo un processo che è inverso rispetto a quello tradizionale, muovendo "dal basso", traendo cioè spunti dalla concretezza della realtà, dai problemi mossi dalla scienza, dalle applicazioni giudiziali, da documenti internazionali». Come chiarito nell'Introduzione a questo lavoro, la scelta è caduta su queste tre posizioni giuridiche, tra le varie, innovative forme di tutela ipotizzate in dottrina (non sempre qualificate come diritti fondamentali) perché esse paiono poter costituire un presidio efficace di fronte al principale rischio che accompagna l'avvento dell'intelligenza artificiale: il sovvertimento della concezione antropocentrica della realtà che caratterizza l'essere umano. Il pericolo connesso a un'eventuale perdita del controllo sulle tecnologie intelligenti, infatti, non è rappresentato da scenari fantascientifici, ma dal non riconoscere più la presenza, il reale meccanismo di funzionamento e i limiti di tali sistemi, finendo per essere incapaci di valutarne e sovvertirne i risultati. Una situazione che realizzerebbe un'inedita forma di reificazione dell'essere umano, ridotto a oggetto destinato a subire le conseguenze di un'innovazione tecnologica che non governa più: un risultato incompatibile con la difesa della centralità dell'individuo alla base dell'intera teorica dei diritti.

A conclusione di questa breve disamina del dibattito in materia di nuovi diritti, è opportuno ricordare un'osservazione proveniente da autorevole dottrina, che invita a procedere con particolare cautela al momento di riconoscere, in un determinato ordinamento, spazio a un nuovo diritto[575]. I diritti fondamentali sono, infatti, un congiunto unitario, le cui parti non possono analizzarsi separatamente dal tutto. In particolare, sono numerosi i casi in cui un diritto, al momento della sua concreta applicazione, porta a restringere determinate prerogative altrui[576]. Ad esempio, le garanzie dell'imputato si pongono in antitesi con le esigenze di tutela della vittima di reato e della collettività generale; la libertà di professare la propria religione spesso contrasta con la libertà di coscienza dei non credenti; i diritti di autodeterminazione nei momenti di inizio e fine vita dialogano in modo inscindibile – e particolarmente intricato – con la protezione della salute. Il conflitto tra diritti è risolto con delicate operazioni di bilanciamento da parte del Legislatore e delle corti, in primo luogo quelle costituzionali[577]. I nuovi diritti vanno inquadrati in questo contesto: il loro riconoscimento e applicazione porterà inevitabilmente al bilanciamento con altri diritti fondamentali. Ecco perché è opportuno evitarne il proliferare incontrollato: la comparsa di nuove fattispecie, se troppo numerose, rischia di comprimere eccessivamente quelle consolidate e storicizzate, alla base dei moderni sistemi democratici.

Questo invito all'attenzione, e al rigore dogmatico e argomentativo al momento di teorizzare nuovi diritti, non è, però, un ostacolo alla loro identificazione, quando il mutamento della società, dei costumi, o delle possibilità offerte dalle nuove tecnologie porti alla comparsa di istanze che appaiono meritevoli di considerazione al più alto livello, magari accompagnate dal riconoscimento in norme primarie e subprimarie. Come già detto, questa terza parte del lavoro sarà dedicata all'approfondimento dell'ipotesi che un fenomeno giuridico di tal genere stia avvenendo in conseguenza della rivoluzione causata dall'avvento dell'intelligenza artificiale.

[575] Cfr. in particolare R. BIN, *Nuovi diritti e vecchie questioni cit.*; *Critica della teoria dei diritti cit.*, p. 55 ss. e R. BIN, P. CHIARELLA, *Critica della teoria dei diritti. Conversazione con Roberto Bin*, in *Ordines – per un sapere interdisciplinare nelle istituzioni europee*, 2, 2018, p. 327 e A. PACE, *Problematica delle libertà costituzionali cit.*, p. 27 ss.

[576] È molto efficace l'espressione utilitizzata da Roberto Bin, nei lavori citati alla nota precedente, per esprimere questo concetto: i diritti sono spesso «a somma zero».

[577] Tra i numerosi studi sul bilanciamento si rimanda, da vari punti di vista e senz'animo di completezza, a V. ITALIA, *Il bilanciamento nelle leggi*, Milano, 2016; R. BIN, *Diritti e argomenti: il bilanciamento degli interessi nella giurisprudenza costituzionale*, Milano, 1992; A. MORRONE, *Il bilanciamento nello stato costituzionale: teoria e prassi delle tecniche di giudizio nei conflitti tra diritti e interessi costituzionali*, Torino, 2014; G. SCACCIA, *Proporzionalità e bilanciamento tra diritti nella giurisprudenza delle corti europee*, in *Rivista AIC*, 3, 2017; G. PINO, *Teoria e pratica del bilanciamento: tra libertà di manifestazione del pensiero e tuteladell'identità personale*, in *Danno e responsabilità*, 6, 2003, p. 577 ss. e ai diversi contributi sul tema raccolti in ai contributi raccolti in G. BRONZINI, R. COSIO (A CURA DI), *Interpretazione conforme, bilanciamento dei diritti e clausole generali*, Milano, 2017.

2. Il diritto di conoscere la natura artificiale di un sistema o interlocutore: teorizzazione, limiti e prime ipotesi di riconoscimento

2.1. Intelligenza artificiale e distinguibilità dall'essere umano: assistenti vocali, chatbot, contenuti deepfake

Il tema della distinguibilità dall'essere umano delle tecnologie basate sull'intelligenza artificiale ha interessato gli esperti fin dagli albori della disciplina. Lo stesso Alan Turing, nel teorizzare il suo noto test, identificò il limite oltre al quale una macchina possa dirsi "pensante" nella capacità di comunicare in forma scritta con un essere umano senza essere riconosciuta[578]. Come evidenziato nella prima parte del lavoro, lo sviluppo di sistemi in grado di imitare perfettamente l'essere umano non è stato, per lunghi periodi, l'obiettivo principale della ricerca nel campo dell'intelligenza artificiale. In primo luogo, per la limitata rilevanza pratica della questione: è parso più funzionale cercare di sviluppare macchine razionali, versatili, efficaci e autonome, a prescindere dalla loro somiglianza con l'umano[579]. In secondo luogo, perché il test di Turing – l'interesse verso il quale è stato, in ogni caso, sempre piuttosto elevato[580] – si è dimostrato particolarmente difficile da superare. Ancora oggi, a più di 70 anni dalla sua formulazione, non esistono tecnologie in grado di mascherare per lunghi periodi la loro natura senza contare sulla distrazione degli esseri umani con cui si interfacciano, o sul vantaggio di svolgere operazioni estremamente specifiche in ambiti circoscritti. Deve evidenziarsi, però, che l'impiego comune dell'intelligenza artificiale non è quello del test di Turing, in cui il sistema è utilizzato al solo scopo di verificarne le capacità decettive, ma sono, appunto, tali utilizzi specifici, molti dei quali ormai parte della vita quotidiana.

Nonostante il test rimanga insuperato, dunque, non si possono sottovalutare gli effetti di quelle tecnologie che, per essere riconosciute come tali, richiedano uno sforzo e un'attenzione del tutto irragionevoli nel contesto pratico di riferimento, finendo per confondere gran parte dei loro utenti. Si tratta, prima di tutto, di sistemi utilizzati per automatizzare comunicazioni di breve o media lunghezza, e in particolare *chatbot* e assistenti vocali[581]. Peraltro, la ricerca in materia pare avere

[578] Il riferimento è di nuovo al noto articolo A. TURING, *Computing machinery and intelligence*, in *Mind*, 236, 1950, p. 433 ss., già menzionato nella prima parte del lavoro.

[579] Cfr. S. RUSSELL, P. NORVIG, *Artificial intelligence cit. (4ª ed.)*, p. 2.

[580] Tanto che, come già evidenziato (cfr. n. 18), dal 1990 esiste un premio annuale, il *LoebnerPrize*, per chi si avvicina di più al suo superamento v. J. WAKEFIELD, *The hobbyistscompeting to make AI human*, in BBC news (online), 13 settembre2019.

[581] Cfr. ad es. i risultati raggiunti da D. ADIWARDANA ET AL., *Towards a Human-like Open-Domain Chatbot*, 2020, arXiv:2001.09977; L. ZHOU, J. GAO, D. LI, H. SHUM, *The Design and Implementation of XiaoIce, an Empathetic Social Chatbot*, 2022, http://arxiv.org/abs/1812.08989; V. SERBAN, C. SANKAR, M. GERMAIN, S. ZHANG, Z. LIN, S. SUBRAMANIAN, ET AL., *A Deep Reinforcement Learning Chatbot*, 2017, http://arxiv.org/abs/1709.02349; J. HILL, W. RANDOLPH FORD, I. G. FARRERAS, *Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations*, in *Computers in Human Behavior*, 2, 2015, p. 245-250. Per le modalità di verifica della distinguibilità di *chatbot* e assistenti vocali, v. invece A. VENKATESH ET AL.,*On*

notevoli margini di sviluppo, ed è ragionevole prevedere che tecnologie di questo genere si diffonderanno su larga scala in un futuro prossimo. Ciò porterà all'aumento, in parallelo, dei possibili problemi ad esse connessi. Da questo punto di vista, non può non evidenziarsi come si siano già realizzati scenari inquietanti: si pensi al massiccio utilizzo di *bot* sui social media, al fine di diffondere notizie false e inquinare il dibattito politico, e alle ricadute sull'effettività della libertà di manifestazione del pensiero e, in generale, sulla stessa tenuta del sistema democratico, analizzate nella parte precedente di questo lavoro. Le capacità decettive di determinati sistemi basati sull'intelligenza artificiale, quindi, meritano la massima considerazione da parte del diritto.

Lo scenario è ulteriormente complicato dai possibili vantaggi di nascondere la natura artificiale di un sistema, resi sempre più evidenti dal progredire della ricerca sulle implicazioni psicologiche e sociali dell'interazione tra uomo e macchina.È molto noto il fenomeno della c.d. *uncanny valley*, la cui prima teorizzazione si deve allo studioso di robotica giapponese Masahiro Mori, nel 1970[582]. Il gradimento e l'empatia verso le tecnologie intelligenti crescono di pari passo alla loro verosimiglianza con l'essere umano, ma si arrestano bruscamente quando quest'ultima raggiunga livelli notevoli senza essere perfetta. Robot antropomorfi, protesi e altri sistemi che imitano quasi del tutto il loro equivalente in carne ed ossa, ma rimangono percepibili come artificiali, suscitano, nella maggioranza dei casi, rifiuto e turbamento. L'effetto, ovviamente, scompare qualora la differenza non sia più percepibile: la quota di empatia, in tal caso, è equivalente a quella generata dall'interazione con qualunque altro essere umano. Da questo punto di vista, non è chiaro quale sarebbe l'effetto di dichiarare spontaneamente la natura artificiale di una tecnologia la cui verosimiglianza sarebbe, in astratto, in grado di confondere larga parte degli utenti. È ipotizzabile, però, che l'acquisita consapevolezza di artificialità farebbe precipitare il livello di gradimento nell'*uncanny valley*, generando il menzionato rigetto emotivo suscitato dalle tecnologie percepite come "troppo simili" all'essere umano.

Anche la ricerca sull'utilizzo di chatbot e assistenti vocali in ambito commerciale fa emergere sempre più indizi dell'impatto negativo della *disclosure* della tecnologia sull'efficacia dell'interazione uomo-macchina. Uno studio del 2019, ad esempio, ha messo in luce come l'utilizzo di chatbot nella vendita a distanza porti a risultati paragonabili a quelli di venditori esperti, quando la loro natura artificiale non è rivelata. Se, invece, la presenza della tecnologia è dichiarata all'inizio della comunicazione coi potenziali clienti, l'efficacia cala bruscamente, portando a una diminuzione

Evaluating and Comparing Open Domain Dialog, 2018, http://arxiv.org/abs/1801.03625; J. P. MCINTIRE, L. K. MCINTIRE, P. R. HAVIG, *Methods for chatbot detection in distributed text-based communications*, in *International Symposium on Collaborative Technologies and Systems*, 2010, 463 ss.

[582]M. MORI, *The uncanny valley*, in *IEEE Robotics & Automation Magazine*, 2, 2012, p. 98 ss. (originale *Bukimi no tani*, in *Energy*, 4, 1970, 33 ss.).; per una rivisitazione del problema in termini più moderni, cfr. T. GELLER, *Overcoming the Uncanny Valley*, in *IEEE Computer Graphics and Applications*, 28, 4, p. 11-17, 2008, doi:10.1109/MCG.2008.79.

delle vendite del 79%. Non può trascurarsi, quindi, che la *disclosure* dell'utilizzo di intelligenza artificiale in una determinata attività potrebbe apparire sconveniente a chi agisca per finalità imprenditoriali o genericamente di persuasione[583].

Oltre al settore dell'automazione di brevi comunicazioni, scritte e orali, esiste una seconda applicazione tecnologica in cui il tema della distinguibilità dell'intelligenza artificiale assume particolare rilievo: la produzione di contenuti c.d. *deepfake*. Con l'espressione *deepfake* si indicano contenuti fotografici, audio o video ritraenti situazioni mai avvenute, elaborati con tecniche diintelligenza artificiale generativa comparse negli ultimi anni[584]. Manipolando immagini e altri materiali è possibile, in particolare, produrre contenuti in cui persone reali siano ritratte in contesti a cui non hanno preso parte, con un livello di precisione tale da rendere pressoché impossibile individuare l'intervento della tecnologia con l'occhio umano. Alcune vicende, già accadute, hanno assunto particolare notorietà: video pornografici artificiali di note personalità vengono ciclicamente diffusi sul web[585]; nel 2020, un video manipolato della politica americana Nancy Pelosi fu utilizzato dai suoi avversari, nel tentativo di farla sembrare ubriaca[586]; un falso video di Mark Zuckerberg intento a pronunciare dichiarazioni compromettenti fu diffuso pochi giorni dopo[587]. Questi primi episodi hanno messo in luce i possibili rischi connessi a un futuro in cui i contenuti*deepfake* venissero normalizzati. Infatti, materiali manipolati potrebbero essere utilizzati con finalità ricattatorie o diffamatorie (paradigmatico è il caso della diffusione di pornografia prodotta con l'intelligenza artificiale) con effetti irreparabili per la personalità morale delle persone coinvolte[588].Né possono trascurarsi i rischi di natura superindividuale:*deepfake*, infatti, potrebbero essere utilizzati al fine di confondere l'opinione pubblica e inquinare il dibattito democratico, fino ad arrivare a un mondo in cui risulti quasi impossibile distinguere con certezza quali contenuti diffusi a scopo d'informazione o propaganda siano veritieri e quali alterati[589].

---

[583]X. LUO ET AL., *Machines vs humans: the impact of artificial intelligence chatbot disclosure on customer purchases*, in *Marketing science*, 6, 2019, p. 937 ss.

[584] Cfr. ad es. *Deepfake*, in *Techopedia*, https://www.techopedia.com/definition/33835/deepfake (22 luglio 2022). Per alcune delle tecniche di intelligenza artificiale più utilizzate per la generazione di *deepfake*, v. T. SHEN, R. LIU, J. BAI, Z. LI, *"deep fakes" using generative adversarial networks (gan)*, 2018; D. YADAV, S. SALMANI, *Deepfake: A Survey on Facial Forgery Technique Using Generative Adversarial Network*, in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 2019, p. 852-857.

[585]Sul tema cfr. K. HAO, *Deepfake porn is ruining women's lives. Now the law is finally banning it*, in *MIT Technology Review*, 12 febbraio 2021; *Porn sites won't take down nonconsensual deepfakes*, Wired, 30 agosto 2020, https://bit.ly/3Kdfe0o (23 luglio 2022)

[586]R. RINI, *Deepfakes are coming. We can no longer believe what we see*, The New York Times, 10 giugno 2019.

[587]R. MERZ, D. O'SULLIVAN, *A deepfake video of Mark Zuckerberg presents a new challenge for Facebook*, CNN Business, 12 giugno 2019, https://edition.cnn.com/2019/06/11/tech/zuckerberg-deepfake/index.html (23 luglio 2022).

[588] Cfr. D. K. CITRON, R. CHESNEY, *Deep fakes: a looming challenge for privacy, democracy, and national security*, in *California Law Review*, 107, 2019, p. 1753 ss.

[589]Cfr. ad esempio L. WILKERSON, *Still waters run deep(fakes): the rising concerns of "deepfake" technology and its influence on democracy and the first amendment*, in *Missouri Law Review*, 86, 1, 2021, p. 407-432; M. BRKAN, *Artificial Intelligence and Democracy: The Impact of Disinformation, Social Bots and Political Targeting*, in *Delphi - Interdisciplinary Review of Emerging Technologies*, 2, 2019, p. 66-71; V. DAN, B. PARIS, J. DONOVAN, M. HAMELEERS,

Di fronte a questi pericoli, e agli altri che potrebbero sorgere nel prossimo futuro, la previsione di un diritto a conoscere se ci si trovi di fronte a tecnologie intelligenti – e dunque di un obbligo generalizzato, in parallelo, a dichiarare il loro impiego – potrebbe rivelarsi un esempio di regolazione efficace e lungimirante. Elevare tale nuova posizione giuridica al rango di vero e proprio diritto fondamentale pare giustificato in virtù dei valori di fondo coinvolti. Il diritto alla *disclosure* dell'intelligenza artificiale sarebbe posto a presidio della sfera della personalità individuale, e in primo luogo delle possibilità di autocoscienza e autodeterminazione, minacciate dal rischio di un'inedita reificazione dell'essere umano, sottoposto al dominio di tecnologie delle quali neppure percepirebbe l'esistenza. Si tratta, del resto, delle stesse preoccupazioni che hanno portato, nei decenni passati, al riconoscimento di nuovidiritti. Ad esempio, le vicende dei diritti afferenti alla sfera dell'identità, analizzate nella parte precedente del lavoro, rappresentano un caso paradigmatico di come gli avanzamenti della tecnicapossano portare alla necessità di riconoscere il massimo livello di tutela a nuove situazioni giuridiche, al fine di mantenere la centralità dell'essere umano.[590]

2.2. I primi esempi di positivizzazione in alcuni ordinamenti e le prospettive aperte nell'Unione Europea dalla Proposta di Regolamento sull'intelligenza artificiale

A conferma della crescente importanza del fenomeno, obblighi di *disclosure* dell'impiego di intelligenza artificiale sono già previsti in alcuni ordinamenti. Uno degli esempi più significativi è il *Bolstering Online Transparency Act* approvato nell'ottobre del 2018 dal Parlamento della California, pienamente in vigore dal luglio 2019[591]. Tale legge proibisce l'utilizzo di *bot* online per finalità di vendita di beni e servizi o propaganda elettorale, qualora non ne sia dichiarata la natura artificiale[592]. Il *BOT Act*, quindi, riconosce, nel circoscritto ambito di riferimento, un obbligo di

---

J. ROOZENBEEK, S. VAN DER LINDEN ET AL., *Visual Mis- and Disinformation, Social Media, and Democracy*, in*Journalism & Mass Communication Quarterly*, 98, 3, 2021, p. 641–664; D. K. CITRON, R. CHESNEY, *Deep fakes: a looming challenge for privacy, democracy, and national security cit*.

[590] Sul tema si rimanda, *ex multis*, ai già citati S. RODOTÀ, *Elaboratori elettronici e controllo sociale*, Bologna, 1973 e *Tecnologie e diritti*, Bologna, 1995; G. PINO, *Il diritto all'identità personale ieri e oggi. Informazione, mercato, dati personali*, in R. PANETTA (A CURA DI), *Libera circolazione e protezione dei dati personali*, 2006, p. 257-321. È doveroso, inoltre, il richiamo alla disciplina giuridica che ha preso il nome di biodiritto, riuscendo a sistematizzare e ricondurre a unità le situazioni generate dall'innovazione tecnologica e dall'evoluzione di cultura e società che, pur nella loro estrema diversità, ponevano sempre questioni inedite e urgenti per la tutela dei diritti fondamentali, in primo luogo la vita e l'integrità fisica. Cfr. ad esempio S. RODOTÀ, P. ZATTI (DIRETTO DA), *Trattato di biodiritto*, Milano, 2010; C. CASONATO, *Introduzione al biodiritto*, Torino, 2012.

[591] California Senate Bill 1001, *Bolstering Online Transaparency Act*, 28 settembre 2018, per alcuni commenti, anche in chiave critica, cfr. B. STRICKE, *People v. Robots: A Roadmap for Enforcing California's New Online Bot Disclosure Act*, in *Vanderbilt Journal of Entertainment and Technology Law*, 4, 4, 6, 2020, p. 839 ss.; J. F. WEAVER, *We Need the California Bot Bill, but We Need It to Be Better*, in *RAIL: The Journal of Robotics, Artificial Intelligence & Law*, 1, 6, 2018, p. 431 ss.; M. HINES, *I Smell a Bot: California's S.B. 1001, Free Speech, and the Future of Bot Regulation*, in *Houston Law Review*, 57, 2, 2019, p. 40 ss.

[592] Nello specifico, il *B.O.T. Act* aggiunge il Chapter 6 alla Part 3 della Division 7 del *Business and Professions Code* californiano, il cui secondo paragrafo (n. 17941, nella numerazione completa) recita: «(a) It shall be unlawful for any

trasparenza dell'impiego di intelligenza artificiale completo e generalizzato. Ciò nonostante, la norma sembra presentare alcuni elementi di debolezza, in grado di minarne l'efficacia. In primo luogo, a presidio dell'obbligo sono poste solo sanzioni amministrative comminate dai poteri pubblici, ela persona fisica che venisse ingannata da un *bot* proibitonon ha mezzi di tutela specifici. In secondo luogo, in capo a *social network* e *internet service provider*, in ossequio al principio dell'esenzione da responsabilità per gli intermediari, non sono posti obblighi di controllo, né obblighi di esclusione dai loro servizi di quegli operatori che venissero, di volta in volta, individuati e sanzionati per occultare l'uso di strumenti di comunicazione automatizzata[593]. Un'eventuale scelta differente avrebbe potuto avere un impatto considerevole, posto che in California si trova la sede legale delle principali piattaforme statunitensi.

Altri esempi di parziale positivizzazione del diritto in esame si rinvengononella già menzionata *Directive on automated decision-making* canadese e nel GDPR europeo, che impongono oneri di *disclosure* dell'intelligenza artificiale qualora sia impiegata per l'automazione, totale o parziale, di un processo decisionale. In particolare, il paragrafo 6.2.1 della direttiva canadese obbliga i poteri pubblici a «providing notice through all service delivery channels in use that the decision rendered will be undertaken in whole or in part by an Automated Decision System»[594]. Invece, gli artt. 13.2 lett. f), 14.2 lett. g) e 15.1 lett. h) del Regolamento europeo, che disciplinano gli obblighi

---

person to use a bot to communicate or interact with another person in California online, with the intent to mislead the other person about its artificial identity for the purpose of knowingly deceiving the person about the content of the communication in order to incentivize a purchase or sale of goods or services in a commercial transaction or to influence a vote in an election. A person using a bot shall not be liable under this section if the person discloses that it is a bot.

(b) The disclosure required by this section shall be clear, conspicuous, and reasonably designed to inform persons with whom the bot communicates or interacts that it is a bot».

[593] Il paragrafo 17942 lett. c) dell'appena citato *Business and Professions Code*, anch'esso aggiunto dal *B.O.T. Act*, recita: «This chapter does not impose a duty on service providers of online platforms, including, but not limited to, Web hosting and Internet service providers». Sottolinea queste criticità, in particolare, J. F. WEAVER, *We Need the California Bot Bill, but We Need It to Be Better cit.*

[594] Ai sensi dell'Appendix C della *Directive*, la tipologia di informazioni richiesta varia a seconda del livello di rischio del sistema automatizzato coinvolto. In caso di rischio moderato o alto, è previsto che la comunicazione sulla presenza di una tecnologia intelligente avvenga tramite «plain language notice posted through all service delivery channels in use (Internet, in person, mail or telephone)». In caso di rischio considerato molto alto, sono previsti adempimenti supplementari: «Plain language notice through all service delivery channels in use (Internet, in person, mail or telephone). In addition, publish documentation on relevant websites about the automated decision system, in plain language, describing: how the components work; how it supports the administrative decision; results of any reviews or audits; and a description of the training data, or a link to the anonymized training data if this data is publicly available». La divisione delle tecnologie di decisione automatizzata in distinti livelli di rischioè disposta dall'Appendix B della direttiva, rubricato *Impact Assessment Leves* (cui l'Appendix C – rubricato *Impact Level Requirements* - rinvia). Il livello di rischio è determinato in base al pericolo che si ritiene il sistema rappresenti per alcuni beni essenziali: «rights of individuals or communities; the health or well-being of individuals or communities; the economic interests of individuals, entities, or communities; the ongoing sustainability of an ecosystem».Nello specifico, sono identificate quattro classi di rischio: minimo (*little or no impact* - Level I), con la specificazione che «Level I decisions will often lead to impacts that are reversible and brief»; moderato (*moderate impacts* - Level II), le cui decisioni «will often lead to impacts that are likely reversible and short-term»; alto (*high impacts* - Level III), col relativo chiarimento «Level III decisions will often lead to impacts that can be difficult to reverse, and are ongoing»; molto alto (*very high impacts* - Level IV), relativo all'automazione di decisioni che «will often lead to impacts that are irreversible, and are perpetual».

informativi nei confronti dell'interessato del trattamento, a seconda del contesto in cui i dati sono raccolti, impongono di fornire informazioni «sull'esistenza di un processo decisionale automatizzato, compresa la profilazione»[595]. È agevole identificare il limite principale di ambo le normative: l'ambito di applicazione estremamente limitato. Sia la direttiva canadese che il Regolamento europeo, infatti, impongono di dichiarare l'utilizzo di intelligenza artificiale solo nel caso di decisioni automatizzate (nel caso del GDPR, solo se frutto del trattamento di dati personali), escludendo numerose applicazioni in cui la mancata *disclosure* delle tecnologie avanzate può avere conseguenze significative (si pensi alle ipotesi, già analizzate, dell'uso di *chatbot* al fine di inquinare il dibattito pubblico, o della produzione di contenuti *deep fake*).

Accanto a queste prime regolazioni settoriali[596], emerge la Proposta di Regolamento dell'Unione Europea sull'intelligenza artificiale presentata nell'aprile 2021, chedimostra una buona consapevolezza del problema, proponendo soluzioni di respiro molto più ampio[597]. Ad occuparsene, nello specifico, è l'art. 52 del testo, che impone obblighi supplementari di trasparenza a determinati sistemi di IA, a prescindere dalla fascia di rischio di appartenenza (e, dunque, anche nel caso non siano considerati ad alto rischio, e non vi si applichino i requisiti tecnici stringenti previsti agli artt. 8 ss.)[598]. Il par. 1 del testo stabilisce l'obbligo generalizzato in capo ai fornitori di sistemi di IA

[595] Tutte e tre le norme impongono di fornire all'interessato informazioni sulla «esistenza di un processo decisionale automatizzato, compresa la profilazione, di cui all'articolo 22, paragrafi 1 e 4, e, almeno in tali casi, informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento per l'interessato». Per un commento v. L. GRIECO, *Informazioni e accesso ai dati personali – artt. 13, 14, 15*, in L. BOLOGNINI, E. PELINO (A CURA DI), *Il codice dela disciplina privacy*, Milano, 2019, p. 147-154 e 158-164.

[596] Potrebbe accostarsi a queste norme anche la previsione, introdotta nel 2021 nel *Code de la santé publique* francese con il nuovo art. L. 4001-3, di un dovere di *disclosure* dell'utilizzo di tecnologie intelligenti al paziente nell'attività diagnostica e clinica. Si tratta, però, di situazioni simili solo nelle linee generali: la recente novella francese, infatti, nella grande maggioranza dei casi non riguarda tecnologie il cui funzionamento possa essere confuso con l'attività di un essere umano, ma sistemi il cui impiego possa non essere percepito dal paziente perché avviene in contesti in cui non è presente (come nel caso dell'esame automatizzato di una radiografia) o non è cosciente (si pensi a un sistema di chirurgia robotica avanzata). Sul tema, si rinvia all'analisi svolta *infra*, p. 279 ss.

[597] Le altre norme dell'Unione Europea in materia di digitale, invece, non trattano direttamente il problema; nella già menzionata Proposta di *Digital Service Act* presentata nel 2020 dalla Commissione, ad esempio, il termine *deepfake* non compare mai. Rimane il fatto che gli obblighi di controllo sui contenuti diffusi dagli utenti che il testo normativo introdurrà, qualora approvato nella versione corrente, in capo alle grandi piattaforme si ripercuoteranno anche sui contenuti *deepfake*, qualora essi rientrassero, in tutto o in parte, tra i materiali illegali ai sensi di altre normative, in primo luogo il Regolamento in materia di IA, se anche la relativa Proposta venisse approvata. Per un'analisi dell'interazione tra le varie discipline dell'Unione, già in vigore o in discussione, in materia di *deepfake* cfr. EUROPEAN PARLIAMENTARY RESEARCH SERVICE, *Tackling deepfakes in European policy*, luglio 2021, https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2021)690039 (24 luglio 2022).

[598] L'art. 52 della Proposta di Regolamento (rubricato *Obblighi di trasparenza per determinati sistemi di IA*), nella sua interezza, recita: «1. I fornitori garantiscono che i sistemi di IA destinati a interagire con le persone fisiche siano progettati e sviluppati in modo tale che le persone fisiche siano informate del fatto di stare interagendo con un sistema di IA, a meno che ciò non risulti evidente dalle circostanze e dal contesto di utilizzo. Tale obbligo non si applica ai sistemi di IA autorizzati dalla legge per accertare, prevenire, indagare e perseguire reati, a meno che tali sistemi non siano a disposizione del pubblico per segnalare un reato. 2. Gli utenti di un sistema di riconoscimento delle emozioni o di un sistema di categorizzazione biometrica informano le persone fisiche che vi sono esposte in merito al funzionamento del sistema. Tale obbligo non si applica ai sistemi di IA utilizzati per la categorizzazione biometrica, che sono autorizzati dalla legge per accertare, prevenire e indagare reati.3. Gli utenti di un sistema di IA che genera o manipola immagini o contenuti audio o video che assomigliano notevolmente a persone, oggetti, luoghi o altre entità o

destinati a interagire con persone fisiche di predisporre sistemi che informino queste ultime della presenza della tecnologia, «a meno che ciò non risulti evidente dalle circostanze e dal contesto di utilizzo». Il par. 3 si occupa, invece, nello specifico di contenuti *deep fake*[599], prevedendo che gli utilizzatori di un sistema che «genera o manipola immagini o contenuti audio o video che assomigliano notevolmente a persone, oggetti, luoghi o altre entità o eventi esistenti e che potrebbero apparire falsamente autentici o veritieri per una persona» siano tenuti a «rendere noto che il contenuto è stato generato o manipolato artificialmente». Nell'ottica adottata da questo lavoro, è di particolare interesse notare come la Proposta di Regolamento sembri concepire queste garanzie allo stesso livello dei diritti fondamentali enunciati nella Carta di Nizza e degli altri interessi primari che dei diritti tradizionalmente rappresentano limite e bilanciamento, in primo la sicurezza e l'ordine pubblico. Lo stesso par. 1 dell'art. 52, infatti, chiarisce che l'obbligo di *disclosure* dei sistemi di intelligenza artificiale interattivi non si applica «ai sistemi di IA autorizzati dalla legge per accertare, prevenire, indagare e perseguire reati, a meno che tali sistemi non siano a disposizione del pubblico per segnalare un reato». L'approccio basato sui diritti emerge con ancor maggiore chiarezza dal secondo comma del paragrafo 3 del medesimo articolo, che indica le eccezioni al dovere di esplicitare sempre la natura artificiale dei contenuti *deep fake*. La norma,

---

eventi esistenti e che potrebbero apparire falsamente autentici o veritieri per una persona ("deep fake") sono tenuti a rendere noto che il contenuto è stato generato o manipolato artificialmente. Tuttavia il primo comma non si applica se l'uso è autorizzato dalla legge per accertare, prevenire, indagare e perseguire reati o se è necessario per l'esercizio del diritto alla libertà di espressione e del diritto alla libertà delle arti e delle scienze garantito dalla Carta dei diritti fondamentali dell'UE, e fatte salve le tutele adeguate per i diritti e le libertà dei terzi.4. I paragrafi 1, 2 e 3 lasciano impregiudicati i requisiti e gli obblighi di cui al titolo III del presente regolamento».

[599] Il fenomeno dei contenuti *deepfake*, finora, è stato affrontato dai pochi ordinamenti che l'hanno preso in considerazione con norme sanzionatorie, anche di diritto penale, che puniscono chi crea o diffonde materiali di tal genere, e non con la previsione di obblighi generalizzati di *disclosure* come quello incluso nella Proposta di Regolamento europeo. In particolare, alcuni sistemi giuridici, principalmente tra gli stati federati degli Stati Uniti, hanno equiparato la diffusione di contenuti pornografici *deepfake* a quella di materiali genuini senza il consenso della persona interessata (c.d. *revenge porn*) e hanno proibito l'impiego di *deepfake* riguardanti i candidati a una competizione elettorale. Anche da questo punto di vista, le norme dello stato americano della California si dimostrano particolarmente avanzate. Nel 2019 è stato approvato l'Assembly Bill 602, *Depiction of individual using digital or electronic technology: sexually explicit material: cause of action*, che sanziona chi «[E]ither (1) creates and intentionally discloses sexually explicit material if the person knows or reasonably should have known the depicted individual did not consent to its creation or disclosure or (2) who intentionally discloses sexually explicit material that the person did not create if the person knows the depicted individual did not consent to its creation», riconoscendo al soggetto leso il diritto di agire in giudizio contro di esso. Nello stesso anno, la California ha introdotto l'Assembly Bill 730, *Elections: deceptive audio or visual media*, che sanziona la produzione e diffusione di contenuti *deepfake* al fine di alterare la percezione pubblica di candidati ad elezioni politiche. Una norma simile è stata introdotta anche in Texas con il *Senate Bill* 751 del 2019, *relating to the creation of a criminal offense for fabricating a deceptive video with intent to influence the outcome of an election*. Sempre nel 2019, lo stato della Virginia ha introdotto il § 18.2-386.2 del Virginia Code, *Unlawful Dissemination or Sale of Images of Another; Penalty*, che criminalizza la produzione e diffusione di *deepfake* pornografici. Gli ordinamenti nazionali europei, invece, non hanno, per ora, emanato norme volte a sanzionare specificamente la diffusione di *deepfake*. Per quanto riguarda l'Italia, i limiti imposti dal principio di tassatività in materia penale portano a pensare che il c.d. reato di *revenge porn* (612-ter c.p.) introdotto nel 2019 non sia applicabile al caso di contenuti pornografici artificiali. L'unica norma a punire espressamente i *deepfake* nel nostro ordinamento, allo stato dell'arte, è l'art. 600 *quater*1 c.p., in materia di contenuti pedopornografici artificiali. Per un commento, cfr. N. ORDONSELLI, *"Porno deepfake": profili di diritto penale*, in *CyberLaws*, 18 gennaio 2021, https://www.cyberlaws.it/2021/porno-deep-fake-profili-penalistici-reato/ (24 luglio 2022).

infatti, prevede che l'obbligo non si applichi qualora «l'uso [del sistema di IA con cui è prodotto il *deep fake*] è autorizzato dalla legge per accertare, prevenire, indagare e perseguire reati o se è necessario per l'esercizio del diritto alla libertà di espressione e del diritto alla libertà delle arti e delle scienze garantito dalla Carta dei diritti fondamentali dell'UE, e fatte salve le tutele adeguate per i diritti e le libertà dei terzi».

3. Il diritto a una spiegazione dei risultati di un sistema: teorizzazione, limiti e prime ipotesi riconoscimento

3.1. Machine learning, reti neurali e sistemi c.d. black-box. Il dibattito sull'opacità dell'intelligenza artificiale e il possibile ruolo del diritto

Il tema della c.d. opacità dell'intelligenza artificiale è già emerso in più punti del lavoro, in primo luogo al momento di elencare i principali problemi etici ad essa connessi[600]. In via di massima approssimazione, il termine indica le difficoltà di vario genere che l'essere umano incontra nel comprendere il funzionamento delle tecnologie intelligenti, interpretarne gli output, rinvenire o costruire una giustificazione razionale di questi ultimi. È intuitivo che capirestruttura, strategie di sviluppo e modalità d'impiego degli strumenti basati sull'intelligenza artificiale è, nella vasta maggioranza dei casi, esclusiva di una ristretta cerchia di tecnici, dotati di competenze specialistiche di alto livello. La questione, però, si spinge oltre la semplice complessità tecnica: come subito si dirà, infatti, alcune applicazioni dell'intelligenza artificiale presentano un certo grado di opacitàcome una vera e propria caratteristica strutturale. Ciò significa che non risultano comprensibili – ferme le considerazioni che saranno svolte sull'ambiguità del concetto - nemmeno ai membri più preparati della comunità di esperti di riferimento[601].

Si è già fatta menzione della distinzione tra intelligenza artificiale *model-based* e *data-driven*[602]. La prima affonda le basi nei sistemi esperti degli anni '70 e '80, e adotta, come approccio

[600] Cfr. *(alcuni) dilemmi etici dell'intelligenza artificiale*, *supra* p. 56 ss.

[601] Cfr. ad es., in generale e da vari punti di vista, R. V. YAMPOLSKIY, *Unexplainability and incomprensibility of artificial intelligence*, 2019, arXiv:1907.03869; G. VILONE, E. LONGO, *Explainable artificial intelligence: a systematic review*, 2020, arXiv:2006.00093 e*Notions of explainability and evaluation approaches for explainable artificial intelligence*, in *Information Fusion*, 76, 2021, p. 89–106; F. PASQUALE, *The black-box society: the secret algorithms that control money and information*, Harvard, 2016; A. PAEZ, *The Pragmatic Turn in Explainable Artificial Intelligence (XAI)*, in *Minds and Machines*, 29, 2019, p. 441-459; Y. BATHAEE, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, in *Harvard Journal of Law & Technology*, 31, 2, 2018, p. 889-938; A. RAI, *Explainable AI: from black box to glass box*, in *Journal of the Academy of Marketing Science*, 2020, 48, 1, p. 137–141; A. ADADI, M. BERRADA, *Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)*, in *IEEE Access*, 6, 2018, 5244-5245; T. WISCHMEYER, *Artificial Intelligence and Transparency: Opening the Black Box*, in T. WISCHMEYER, T. RADEMACHER (A CURA DI), *Regulating Artificial Intelligence. Springer*, Cham, 2020 https://doi.org/10.1007/978-3-030-32361-5_4; D. CASTELVECCHI, *Can we open the black box of AI?*, in *Nature*, 2016, 538(7623), p. 20–23.

[602] Si rinvia, tra i molti, a S. RUSSELL, P. NORVIG, *Artificial intelligence cit.*, p. 47 ss.; T. WEI, X. CHEN, X. LI, Q. ZHU, *Model-based and data-driven approaches for building automation and control*, in *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2019, doi:10.1145/3240765.3243485; S. FORMENTIN, K, VON

fondamentale, la programmazione minuziosa delle regole necessarie a svolgere le attività per cui è progettata. La seconda, fondata prima di tutto sull'osservazione di grandi moli di dati, è alla base del settore dell'apprendimento automatico, come già visto caratterizzato da ragguardevoli progressi negli ultimi due decenni. La discussione sull'opacità dell'intelligenza artificiale si concentra su quest'ultima famiglia di tecnologie, e in particolare sulle reti neurali.

Come già analizzato nella prima parte, il funzionamento delle reti neurali si basa sull'elaborazione dell'input ricevuto tra i differenti strati di "nodi" della rete, fino allo strato che comunica l'output all'esterno[603]. L'apprendimento avviene attraverso la memorizzazione e continua correzione del valore che ciascun nodo della rete associa a un determinato input. Una volta "allenato" su un *dataset* apposito, così, il sistema è in grado di operare nel mondo reale, fornendo risultati il cui livello di precisione, spesso, supera di gran lunga quello dell'essere umano, o quello ottenibile con ogni altra tecnologia. Struttura e funzionamento delle reti neurali ricordano, nelle linee essenziali, quanto avviene a livello di connessioni sinaptiche tra i neuroni del cervello umano. Inoltre, in numerosi ambiti applicativi, più la rete è "profonda", ovvero costituita da numerosi strati di neuroni, e più le sue prestazioni risultano efficaci. Ciò ha portato al successo vertiginoso del *deep learning* che ha caratterizzato gli ultimi anni, già trattato nella prima parte[604].

Proprio la profondità della rete neurale pare direttamente correlata alla sua opacità.Infatti, il funzionamento della rete, inteso come il percorso che, dato un determinato input, conduce a un determinato output, ottenuto assegnando un diverso "peso specifico" a ciascuno dei nodi che la compongono, sembra impossibile da comprendere razionalmente. In primo luogo, in ragione delle dimensioni del problema: anche per applicazioni dell'intelligenza artificiale relativamente semplici, la ricostruzione delle connessioni tra i nodi della rete risulterebbe estremamente laboriosa e complessa, e raggiungerebbe un'estensione intrattabile per le abilità dell'essere umano (al pari, del resto, dei collegamenti sinaptici all'interno del cervello)[605]. In secondo luogo, perché tali

---

HEUSDEN, A. KARIMI, *Model-based and data-driven model-reference control: a comparative analysis*, in *2013 European Control Conference (ECC)*, 2013, 10.23919/ECC.2013.6669388; S. A. YABLONSKY, *Multidimensional data-driven artificial intelligence innovation*, in *Technology innovation management review*, 9, 12, 2019, p. 16-28; K. MANHART, *Artificial Intelligence Modelling: Data Driven and Theory Driven Approaches*, in K. G. TROITZSCH (A CURA DI), *Social Science Microsimulation*, Berlino, 1996, p. 416-431.

[603]Cfr., in generale, C.C. AGGARWAL, *Neural networks and deep learning – a textbook*, Berlino, 2018; 1441-1461; L. DENG, D. YU, *Deep Learning: Methods and Applications*, in *Foundations and Trends in Signal Processing*, 7, 3-4, p. 198-205, doi:10.1561/2000000039; E. R. KANDEL, J. H. SCHWARTZ, T. M. JESSELL, S. A. SIEGELBAUM, A. J. HUDSPETH (A CURA DI), *Principles of neural science (Vth ed.)*, New York, p. 1141-1161; J. L. MCCLELLAND, M. BOTVINICK, *Deep learning: Implications for human learning and memory, in The Oxford Handbook of Human Memory*, 2020, doi.10.31234/osf.io/3m5sb; N.J. NILSSON, *The Quest for Artificial Intelligence cit.*, p. 413 ss.

[604] Sul punto, v. ad es. S. RUSSELL, P. NORVIG, *Artificial intelligence cit. (4 ^{a} ed.)*, p. 26-27; A. KRIZHEVSKY, I. SUTSKEVER, G.E. HINTON, *ImageNet classification with deep convolutional neural networks*, in *NIPS*, 1, 2012.

[605] È noto, infatti, che le conoscenze sul funzionamento del cervello, e sul legame tra stati biologici e stati mentali, sono ancora estremamente parziali, nonostante gli importantissimi progressi delle neuroscienze negli ultimi decenni, cfr. ad es. J. FODOR, *How the mind works: what we still don't know*, in *Daedalus*, 2006, 135, 3, p. 86-94; D. POPPEL, W. IDSARDI, *We don't know how the brain stores anything, let alone words*, in *Trends in Cognitive Sciences*, 26, 12, 2022,

informazioni non sono minimamente paragonabili a una motivazione come comunemente intesa: forniscono, semplicemente, il "percorso" con cui la rete neurale giunge a un determinato risultato a partire dal relativo input. La logica ad esso sottesa, però, rimane del tutto inaccessibile, non essendo ricavabile da un'eventuale mappatura dei legami tra i diversi nodi della rete da essa "appresi" in fase di addestramento[606]. Com'è intuitivo, più la rete è profonda, più l'opacità, nel senso appena descritto, risulta marcata, tanto da essere generalmente indicata con l'evocativa espressione *black-box*[607].

Riducendo la questione all'essenziale, è per questa ragione che il problema dell'opacità è associato principalmente al settore del *machine learning* nella letteratura scientifica che si occupa, da vari punti di vista, del possibile impatto sociale dell'intelligenza artificiale[608]. Infatti, se la ragione principale dell'incomprensibilità delle tecnologie riconducibili all'approccio *model-based* è la loro complessità tecnica, quella dell'opacità di molti sistemi di apprendimento automatico è qualitativamente differente. L'acquisizione di competenze specialisticherende comprensibile l'intelligenza artificiale *model-based*, quantomeno ai tecnici che si occupano della programmazione delle regole di funzionamento di volta in volta impiegate. Invece, nemmeno i tecnici del settore e gli sviluppatori che si siano occupati in prima persona dell'addestramento di determinate tecnologie basate sul*machine learning*, spesso,sono in grado di ricavare, osservandone il comportamento, una motivazione dei loro output. In questo caso, il problema non si esprime in termini di complessità, ma di vera e propria impossibilità tecnica, almeno parziale. A questa distinzione di principio devono aggiungersi, però, due importanti specificazioni. Va tenuto presente, in primo luogo, che i diversi approcci all'intelligenza artificiale sono sempre più spesso combinati all'interno di applicazioni tecnologiche molto complesse, che mettono in crisi la suddivisione della disciplina in sottosettori, e,

---

p. 1054-1055. L'attribuzione di una mente simile alla nostra agli altri esseri razionali è, sostanzialmente, frutto di una convenzione sociale e non della conoscenza dei meccanismi biologici che ad essa danno origine, cfr. D.C. DENNET, *The intentional stance*, Cambridge, 1987. Significativi progressi della nostra comprensione del cervello, peraltro, sono derivati dall'intuizione di indagarne il funzionamento a partire da quello delle tecnologie avanzate, con una radicale inversione di prospettiva, cfr. ad. es. D. GRAHAM, *An internet in your head*, NewYork, 2021.

[606] Cfr. ad es. D. WEINBERGER, *Our machines now have knowledge we will never understand*, in *Wired*, 10 aprile 2017, https://bit.ly/3ejszrV (2 agosto 2022); S. QUINTARELLI (A CURA DI), *Intelligenza artificiale. Cos'è davvero, come funziona, che effetti avrà*, Torino, 2020; C. KOZYRKOV, *Explainable Ai won't deliver. HerÈs why*, in *Medium*, 16 novembre 2020, https://bit.ly/3CS5dnK (2 agosto 2022), oltre ai già citati D. CASTELVECCHI, *Can we open the black box of AI? cit.*; R. V. YAMPOLSKIY, *Unexplainability and incomprensibility of artificial intelligence cit.*;FY. BATHAEE, *The Artificial Intelligence Black Box and the Failure of Intent and Causation cit.*; A. ADADI, M. BERRADA, *Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI) cit.*; T. WISCHMEYER, *Artificial Intelligence and Transparency: Opening the Black Box cit.*

[607] Un contributo decisivo a rendere l'espressione d'uso comune, anche fuori dagli ambiti specialistici, è venuto dal successo dell'opera di F. PASQUALE, *The black-box society: the secret algorithms that control money and information cit.*

[608] Cfr. ad es. S. QUINTARELLI, F. COREA, F. FOSSA, A. LOREGGIA, S. SAPIENZA, *AI: Profili etici. Una prospettiva etica sull'intelligenza artificiale: principi, diritti e raccomandazioni cit.*, p. 197-198; o le già più volte citate HIGH LEVEL EXPERT GROUP ON AI, *Ethics Guidelines for Trustworthy Artificial Intelligence cit.*, p. 25.

di conseguenza, le relative ripercussioni in materia di opacità[609]. In secondo luogo, i moderni algoritmi *model-based*, basati sulla programmazione della conoscenza, raggiungono spesso dimensioni tali da poter funzionare solo con l'impiego di computer dotati di altissima capacità computazionale, e l'idea di comprenderne il codice, per un essere umano, risulta del tutto impraticabile[610]. In tali casi, anche se la distinzione concettuale vista poche righe sopra rimane valida, l'effetto pratico è che il sistema risulta nei fatti inaccessibile tanto quanto le reti neurali più profonde.

Alcune voci critiche, al fine di circoscrivere la portata del problema dell'opacità, evidenziano come sia presente, anche nel dibattito specialistico, molta confusione attorno allo stesso concetto di motivazione o spiegazione. In particolare, andrebbero nettamente distinte due attività: interpretare il sistema e spiegarne i risultati[611]. L'interpretazione di una rete neurale consiste nella menzionata opera di decostruzione e mappatura dei legami tra i nodi che la compongono e che hanno portato da un determinato input al relativo output. Un progetto, come già visto, generalmente impraticabile per ragioni di scala, e comunque non utile alla reale comprensione delle conclusioni cui l'algoritmo è giunto. La costruzione di una motivazione logica che connetta dati di partenza e risultato finale del sistema, in grado di soddisfare un utente umano, è, invece, il nocciolo dell'attività di vera e propria spiegazione della tecnologia. In questa accezione, la spiegazione può anche prescindere, in tutto o in parte, da tipologia e modalità di funzionamento del sistema, e consistere, semplicemente, nell'elaborazione *ex-post* di un giustificativo razionale dell'*output* da esso fornito. La difficoltà o impossibilità di costruire tal genere di motivazione sarebbe un indicatore della possibile presenza di un errore nei risultati dell'algoritmo, eventualità non eliminabile, posta la natura statistica delle tecnologie in esame, come già chiarito in più punti di questo lavoro[612]. Coloro che tendono a minimizzare la questione dell'opacità sottolineano che il cervello umano presenterebbe la stessa caratteristica. L'eventuale ricostruzione precisa dei legami sinaptici tra i neuroni, infatti, non

[609] Cfr. ad. es. *Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report*, 2021, https://stanford.io/3KO5R7K (2 agosto 2022).

[610] Sul punto, cfr. in particolare le osservazioni di Z. C. LIPTON, *The mythos of model interpretability*, https://doi.org/10.48550/arXiv.1606.03490, 2017.

[611] La distinzione qui adottata tra *interpretazione* e *spiegazione* è elaborata a partire da studi di lingua inglese che esaminano la varietà delle possibili tipologie di spiegazione ed evidenziano la differenza concettuale che si è scelto di ricondurre ai due termini, cfr. ad es. G. MONTAVON, W. SAMEK, K.R. MÜLLER, *Methods for interpreting and under standing deep neural networks*, in *Digital Signal Processing*, 73, 2017, p. 1-15; G. VILONE, L. LONGO, *Explainable artificial intelligence: a systematic review*, https://arxiv.org/pdf/2006.00093.pdf, 2020 (2 agosto 2022); Z. C. LIPTON, *The mythos of model interpretability cit.*; T. SPEITH, *A review of taxonomies of expleinable artificial intelligence (XAI) methods*, in *2022 ACM Conference on fairness, accountability, and transparency*, p. 2239-2250; C. MESKE, E. BUNDE, J. SCHNEIDER, M. GERSCH, *Explainable artificial intelligence: objectives, stakeholders, and future research opportunities*, in *Information System Management*, 39, 1, p. 53-63, https://doi.org/10.1080/10580530.2020.1849465 (2 agosto 2022).

[612] Si rimanda, nuovamente, all'efficace definizione proposta da S. QUINTARELLI, F. COREA, F. FOSSA, A. LOREGGIA, S. SAPIENZA, *AI: Profili etici. Una prospettiva etica sull'intelligenza artificiale: principi, diritti e raccomandazioni*, in *BioLaw Journal – Rivista di BioDiritto*, p. 195 ss: «motore statistico che produce necessariamente risultati probabilistici».

offrirebbe alcuna indicazione sulla bontà di un determinato ragionamento. Anzi, le modalità con cui l'attività cerebrale genera il funzionamento della mente e della coscienza sono ancora, in larga parte, ignote[613]. La stessa attribuzione agli altri di stati mentali simili ai nostri è frutto di una convenzione sociale, per quanto ben radicata[614]. Così, la motivazione razionale delle decisioni, e in generale l'intera conoscenza, sarebbero del tutto indipendenti dal sostrato biologico da cui derivano, del quale sarebbero un modello elaborato *ex post* attraverso l'interazione tra esseri umani e basato su convenzioni, in primo luogo il linguaggio. A conferma di questa tesi, sono spesso richiamati gli studi di psicologia comportamentista che hanno messo in luce come, in molti casi, una decisione umana si basi su elementi emotivi e irrazionali, e la motivazione comunicata all'esterno sia costruita successivamente[615]. Anche il cervello umano, dunque, è una *black-box*, ma questo non ci ha impedito di vivere serenamente finora. Secondo alcuni, allora, la convivenza con tecnologie che presentino la medesima caratteristica non dovrebbe preoccuparci[616].

Non può negarsi che il dibattito sulle possibili conseguenze dell'opacità dell'intelligenza artificiale, negli ultimi anni molto vivace, abbia visto prendere piede anche posizioni drammatizzanti e chiaramente antitecnologiche[617]. Se tali opinioni catastrofiste non sembrano condivisibili, lo stesso deve dirsi di quelle che, magari partendo dall'osservazione di utili dati di realtà, giungono ad affermare, nella sostanza, che quello della *black-box* sia un falso problema, non meritevole di discussione. È doveroso, in ogni caso, prendere coscienza dell'ambiguità del concetto di spiegazione, e in particolare di come ogni spiegazione non riproduca fedelmente le dinamiche che hanno condotto ad essa, sia dal punto di vista della struttura biologica del cervello, che rimane insondabile, che dal punto di vista del ruolo di elementi emotivi o comunque irrazionali. Come detto, il modello logico con cui identifichiamo e comunichiamo le ragioni di un determinato

[613] Cfr. ancora, *ex multis*, J. FODOR, *How the mind works: what we still don't know cit.*; D. GRAHAM, *An internet in your head cit.*, p.

[614] Il referimento è, prima di tutto, ai già menzionati studi di D.C. DENNET, *The intentional stance cit.*

[615] Il rinvio è, in primo luogo, ai già citati studi di D. KAHNEMAN, A. TWERSKY, *Prospect theory: an analysis of decision under risk*, in *Econometrica*, 47, 1979, p. 263-291; D. KAHNEMAN, A. TWERSKY, P. SLOVIC, *Judgment under uncertainity. Heuristics and biases*, Cambridge, 1982, valsi ai due studiosi il premio Nobel per l'economia nel 2002. Cfr. anche il recente D. KAHNEMAN, O. SIBONY, C. R. SUNSTEIN, *Noise: a flaw in human judgment*, Boston, 2021.

[616] Cfr. ad es. V. PANDE, *Artificial intelligence's "black-box" is nothing to fear*, The New York Times, 25 gennaio 2018; P. MARTIN, *Why you should not worry about AI black box*, 23 ottobre 2017, https://bit.ly/3QdTXVw (2 agosto 2022); T. GREEN, *How to stop fearing black box AI and love the robot-ruled future*, TNW, 8 dicembre 2017, https://bit.ly/3BiyarT (2 agosto 2022); A.J. LONDON, *Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability*, in *Hastings Center Report*, 49, 1, 2019, p. 15-21. Si vedano anche le considerazioni di C. KOZYRKOV, *Explainable Ai won't deliver. HerÈs why cit*.

[617] Cfr. ad es. J. BARRAT, *Our final invention: artificial intelligence and the end of the human era*, New York, 2013. Posizioni pessimistiche sulle possibili conseguenze dello sviluppo di tecnologie intelligenti sempre più complesse sono state espresse anche da figure di primo piano del settore della tecnologia, come lo studioso Stephen Hawking o l'imprenditore E. Musk, che si sono detti estremamente preoccupati in varie dichiarazioni pubbliche, cfr. M.MCFARLAND, *Elon Musk: with artificial intelligence we are summoning to the demon*, The New York Times, 24 ottobre 2014; R. CELLAN-JONES, *Stephen Hawking warns artificial intelligence could end mankind*, BBC News – Technology, 2 dicembre 2014, https://www.bbc.com/news/technology-30290540.

fenomeno, o i motivi a supporto di una nostra decisione od opinione, non è una rappresentazione della realtà di ciò che avviene, dentro di noi, a livello biologico ed emozionale, dalla quale potrebbedel tutto prescindere. La decostruzione del concetto di spiegazione, però, non porta a sminuire l'importanza del problema dell'intelligenza artificiale *black-box*, semmai a inquadrarlo nel giusto contesto.

Per quanto anche il funzionamento del cervello, o le reali ragioni di decisioni e valutazioni individuali o collettive risultino insondabili, infatti, i problemi sollevati dall'opacità della tecnologia non possono essere ignorati. Vi è un'ovvia e profonda differenza tra l'attribuire convenzionalmente agli altri esseri umani una mente e una coscienza analoghe alle nostre, e l'eventualità di farlo con una macchina. Non conoscere nel dettaglio i meccanismi biologici che avvengono nel cervello non ci impedisce di presumere che essi siano simili a ciò che avviene nell'organismo di ogni altro essere umano. Inoltre, il pensiero logico-razionale con cui giustifichiamo le nostre azioni e decisioni può essere un semplice modello costruito *ex-post*, che non ne rappresenta le reali ragioni – irrazionali, emotive, destinate a rimanere incomprensibili – ma ciò non ci impedisce di comunicare tali ragionamenti all'esterno a terzi che li riconoscono come validi e che fanno altrettanto[618]. L'impossibilità di percepire le reali cause dei nostri stati mentali e di quelli altrui non porta all'incomunicabilità, né alla radicale sfiducia verso il prossimo. A giocare un ruolo decisivo, però, è la percezione di similitudine: siamo disposti ad attribuire agli altri una mente e una coscienza analoghe alle nostre perché li identifichiamo come appartenenti alla stessa specie[619]. Allo stesso modo, la consapevolezza che il cervello altrui è, di fatto, una *black-box* non ci turba più di quanto lo faccia apprendere che lo è anche il nostro, perché ci percepiamo membri della stessa comunità. Il ragionamento, però, non può estendersi alle macchine, la cui alterità, invece, ci appare radicale[620].

[618]Pare doveroso, sul punto, un riferimento agli studi sulla reale natura del fenomeno della comprensione di testi e concetti condotti da alcuni dei più grandi filosofi del '900, che hanno evidenziato come l'acquisizione di nuove conoscenze sul mondo esterno non sia mai "pura", ma sempre il frutto di una complessa interazione tra elementi assunti *ex novo* e componenti pregresse di natura culturale, tradizionale e, forse, innata. Il rimando è inanzitutto ad alcune opere di Martin Heidegger, in particolare *Essere e tempo*, Halle, 1927, e agli studi di Hans-Georg Gadamer (per i quali le tesi di Heidegger hanno costituito le basi fondamentali) in primo luogo *Verità e metodo*, Tubinga, 1960.

[619] Gli studi sulle modalità di sviluppo della c.d. "teoria della mente" (per l'appunto, l'attribuzione di stati emotivi e mentali agli altri) infatti, identificano nell'interazione con adulti e coetanei nella prima infanzia, attraverso il linguaggio, l'osservazione e l'imitazione, il momento cruciale del suo sviluppo, cfr. ad es. A. WHITEN (A CURA DI), *Natural Theories of Mind: Evolution, Development, and Simulation of Everyday Mindreading*, Oxford, 1991; P. CARRUTHERS, P. K. SMITH (A CURA DI), *Theories of theories of mind*, Cambridge, 1996; J. S. BRUNER, *Intention in the structure of action and interaction*, in L. P. LIPSITT, C. K. ROVEE-COLLIER (A CURA DI), *Advances in infancy research,* 1, p. 41–56, Norwood, 1996; D. PREMACK, G. WOODRUF, *Does the chimpanzee have a theory of mind*?, in *Behavioral and brain sciences*, 1978, 1, 4, p. 515-526.

[620] Cfr. B. ERB, *Artificial intelligence & the theory of mind*, in *Seminar Cognitition & Emotion V*, 2016, https://bit.ly/3CVhMyE (3 agosto 2022); T. ARAUJO, *In AI we trust? Perceptions about automated decision-making by artificial intelligence*, in *AI & Society*, 35, 2020, p. 611-653. Un ulteriore indice della difficoltà a percepire la macchina come un proprio simile proviene dalle evidenze scientifiche del gradimento basso o nullo dell'interazione consapevole con tecnologie antropomorfe estremamente realistiche, come il citato fenomeno dell'*uncanny valley*, cfr. M. MORI, *The uncanny valley cit*; T. GELLER, *Overcoming the Uncanny Valley cit*.

La serena accettazione dell'opacità degli altri esseri umani non muta i termini del problema dell'opacità della tecnologia: la prima, infatti, non è comunemente percepita come pericolosa o problematica solamente perché riguarda, appunto, gli esseri umani. L'interfacciarsi con un sistema del quale non è comprensibile il funzionamento pone questioni, come si dirà, potenzialmente in grado di sovvertire radicalmente il rapporto tra l'essere umano e le macchine che produce, riducendo in modo inedito le possibilità di controllo del primo sulle seconde[621]. L'utilizzo di algoritmi *black-box* per scelte e decisioni, inoltre, mette in discussione la stessa possibilità di imputare queste ultime a un essere umano, e quindi ricondurle alla menzionata irrazionalità che siamo disposti a tollerare perché riguarda un nostro simile. Le preoccupazioni sollevate dall'opacità di alcuni sistemi di intelligenza artificiale, dunque, non vengono cancellate dalla consapevolezza che anche l'essere umano, spesso, si rivela una *black-box*, poiché riguardano il ruolo che quest'ultimo attribuisce a sé stesso nel mondo e il modo in cui percepisce la sua condizione, che si fonda sulla convenzione che ogni persona abbia stati mentali ed emotivi paragonabili, che la rendono capace di interpretare razionalmente il mondo esterno, macchine comprese.

Inoltre, nonostante la differenza tra tecnologie *model-based* e intelligenza artificiale basata sul *machine learning* sia stata, per quanto riguarda la comprensibilità dei sistemi, parzialmente ridimensionata dall'aumento della complessità delle prime, preme rilevare che la profonda distinzione concettuale che le caratterizza permane inalterata. Come già analizzato, infatti, l'intelligenza artificiale *model-based* si basa, in termini molto generali, sulla programmazione delle regole con cui dovrà essere portata a termine una determinata attività[622]. In tal modo, viene codificato un modello razionale, analogo a quelli con cui l'essere umano procede nei sui ragionamenti, spiega le sue scelte, motiva le sue azioni, al netto degli studi, già citati, che hanno messo in luce come questi siano, in molti casi, giustificazioni postume di procedimenti irrazionali. Anche se le dimensioni concrete dell'algoritmo rendono spesso irrealistico ricostruirne la struttura, dunque, esso è elaborato imitando il pensiero razionale. Le reti neurali, e in particolare il *deep learning*, funzionano, invece, su base statistica, attraverso l'osservazione di dati. Mentre l'intelligenza artificiale *model-based* imita il modello di ragionamento degli esseri umani, esse si ispirano direttamente al cervello umano, il cui funzionamento, come più volte ripetuto, rimane in buona parte sconosciuto anche dal punto di vista biologico[623]. Le reti neurali, dunque, somigliano a

[621] Sul rapporto tra spiegabilità e controllo umano cfr. in particolare *infra*, p. 221 ss.

[622] Cfr. F. HAYES-ROTH, *Rule-based systems*, in *Communications of the ACM*, 28, 9, 1985 p. 921–932, https://doi.org/10.1145/4284.4286; oltre ai già citati S. RUSSELL, P. NORVIG, *Artificial intelligence cit.*, p. 47 ss.; T. WEI, X. CHEN, X. LI, Q. ZHU, *Model-based and data-driven approaches cit.*; S. FORMENTIN, K, VON HEUSDEN, A. KARIMI, *Model-based and data-driven model-reference control cit.*

[623] Si rinvia ai già citati C.C. AGGARWAL, *Neural networks and deep learning cit.*; E. R. KANDEL, J. H. SCHWARTZ, T. M. JESSELL, S. A. SIEGELBAUM, A. J. HUDSPETH (A CURA DI), *Principles of neural science cit.*; J. L. MCCLELLAND, M. BOTVINICK, *Deep learning: Implications for human learning and memory cit.*

legami sinaptici che non comprendiamo appieno, e la cui eventuale mappatura non apparirebbe, di per sé, collegata col pensiero che riteniamo essi generino; l'intelligenza artificiale *model-based*, invece, imita il modello, elaborato secondo i canoni di ciò che chiamiamo razionalità, con cui organizziamo e comunichiamo quel pensiero. Per questa ragione, al netto dell'ibridazione tra le due categorie, e delle numerose specificità potenzialmente presenti in ogni caso concreto, la differenza tra i due approcci, in termini di comprensibilità, continua ad essere decisiva.

Gli argomenti più utilizzati da parte di chi minimizza il problema, dunque, non convincono. Essi, semmai, possono essere utili per inquadrare i limiti e le caratteristiche delle possibili soluzioni, e il ruolo del diritto nell'indirizzare verso di esse. La consapevolezza che ogni spiegazione razionale consiste in un modello contenente un inevitabile grado di semplificazione, talvolta del tutto indipendente dalla realtà da cui ha origine[624], orienta il ragionamento sui possibili rimedi all'opacità dell'intelligenza artificiale verso alcune direzioni principali. In primo luogo, il problema non è risovibile in termini netti, per cui una tecnologia è spiegabile o no. Ogni possibile spiegazione è destinata a rimanere una parziale giustificazione del comportamento del sistema, fondata, come appena visto, su un'inevitabile approssimazione. La valutazione della comprensibilità o spiegabilità di un sistema, dunque, deve intendersi come un discorso sul livello di opacità tollerabile e sul tipo di spiegazione utile a soddisfare le esigenze dell'essere umano, soglie che spetta – e spetterà sempre di più – fissare anche al diritto. In secondo luogo, le strategie per costruire una possibile spiegazione o motivazione possono essere le più varie, posto che nessun modello è in grado di riprodurre esattamente la realtà di riferimento. Alcune spiegazioni, così, potrebbero ricavarsi in modo del tutto indipendente dalle caratteristiche della tecnologia di riferimento, altre, invece, potrebbero consistere in un tentativo, almeno parziale, di interpretare il funzionamento del sistema, e costruire un ragionamento razionale dai dati acquisiti in tal modo. Si tratta di una varietà di approcci di cui il campo di ricerca dell'*explainable artificial intelligence*, in continua espansione, offre, come subito si dirà, numerosi interessanti esempi. Determinare quali strategie di spiegazione siano preferibili, a seconda del contesto, sarà compito degli specialisti del settore, di chi affronta, da vari punti di vista, il tema dell'impatto sociale dell'intelligenza artificiale, e, non da ultimo, degli studiosi di diritto chiamati a orientare la regolazione di queste tecnologie.

[624] Cfr. T. MILLER, *Explanation in artificial intelligence: insights from the social sciences*, in *Artificial intelligence*, 267, 2019, p. 1-38. Il tema, ovviamente, trascende l'ambito dell'intelligenza artificiale per sconfinare nell'epistemologia, attinendo, in ultima analisi, al rapporto tra ragionamento, linguaggio e realtà. Si vedano, in ottica estremamente generale, J. C. PITT (A CURA DI), *Theories of explanation*, Oxford, 1988; W. G. LYCAN, *Explanation and epistemology*, in P. K. MOSER (A CURA DI), *The Oxford Handbook of epistemology*, Oxford, 2002, p. 408-433.

3.2. Cenni tecnici sull'explainable artificial intelligence: stato dell'arte, limiti e prospettive future

L'appena mezionato campo di studi della c.d. *explainable artificial intelligence*, spesso indicato con la sigla XAI, concentra i sui sforzi sullo sviluppo di tecnologie basate sull'intelligenza artificiale che arginino l'appena analizzato problema dell'opacità[625]. Deve evidenziarsi, prima di tutto, che questo filone di ricerca sta vivendo un vertiginoso picco d'interesse, al pari del dibattito interdisciplinare sulle possibli conseguenze dell'utilizzo di sistemi *black-box*, del quale si sono già tracciate linee essenziali. Ciò rende particolarmente complesso ottenere un quadro generale degli avanzamenti nel campo, anche agli stessi tecnici che vi lavorano. La grande diversità degli approcci adottati dalla ricerca è ben rappresentata dalla circostanza che, nel solo 2021, possono contarsi sei diversi tentativi, condotti da esperti del settore di rilevanza internazionale, di organizzare le iniziative riconducibili all'*explainable artificial intelligence* in una tassonomia unitaria[626]. Peraltro,

[625] Sul tema possono consultarsi, *ex multis*, D. GUNNING, M. STEFIK, J. CHOI, T. MILLER, S. STUMPF, G. Z. YANG, *XAI - Explainable artificial intelligence*, in *Science Robotics*, 4, 37, 2019, https://doi.org/10.1126/scirobotics.aay7120 (2 agosto 2022); A. DAS, P. RAD, *Opportunities and Challenges in Explainable Artificial Intelligence (XAI)*, https://arxiv.org/abs/2006.11371, 2020; F. K. DOŠILOVIĆ, M. BRČIĆ, N. HLUPIĆ, *Explainable artificial intelligence: a survey*, in *41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2018, p. 210-215; W. SAMEK, K.R. MÜLLER, *Towards Explainable Artificial Intelligence*, in W. SAMEK, G. MONTAVON, A. VEDALDI, L. HANSEN, K. R. MÜLLER (A CURA DI), *Explainable AI: interpreting, explaining and visualizing deep learning*, Cham, 2019; R. CONFALONIERI, L. COBA, B. WAGNER, T.R. BESOLD, *A historical perspective of explainable Artificial Intelligence,* in *WIREs Data Mining and Knowledge Discovery*, 1, 2021, https://bit.ly/3cLtL7e (2 agosto 2022); A. BARREDO ARRIETA, N. DÍAZ-RODRÍGUEZ, J. DEL SER, A. BENNETOT, S. TABIK, A. BARBADO ET AL., *Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI*, in *Information Fusion*, 58, 2020, p. 82–115; M.R. ISLAM, M.U. AHMED, S. BARUA, S. BEGUM, *A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks*, in *Applied Sciences*, 12, 3, 2022, p. 1353 ss.; PEDRESCHI D., GIANOTTI F., GUIDOTTI R., MONREALE A., RUGGIERI S., TURINI F., *Meaningful explanations of Black Box AI decision systems*, in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 10.1609/aaai.v33i01.33019780; GUIDOTTI, R. MONREALE A., PEDRESCHI D., *The AI black-box explanation problem*, in *ERCIM News*, 116, 2019, p. 12-13; GUIDOTTI R., MONREALE A., RUGGIERI S., TURINI F., GIANOTTI F., PEDRESCHI D., *A survey of methods for explaining black box models*, in *ACM Computing Surveys*, 51, 5, 2018, p. 93 ss.; A. KRAJNA, M. KOVAC, M. BRCIC, A. SARCEVIC, *Explainable artificial intelligence: an updated perspective*, in *45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*, 2022, https://bit.ly/3RdFqe4 (2 agosto 2022); D. DORAN, S. SCHULZ, T. R. BESOLD, *What Does Explainable AI Really Mean? A New Conceptualization of Perspectives*, 2017, http://arxiv.org/abs/1710.00794 (2 settembre 2022), G. VILONE, E. LONGO, *Explainable artificial intelligence: a systematic review cit.* e *Notions of explainability and evaluation approaches for explainable artificial intelligence cit.*; A. PAEZ, *The Pragmatic Turn in Explainable Artificial Intelligence cit.*; A. RAI, *Explainable AI: from black box to glass box cit.*; A. ADADI, M. BERRADA, *Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI) cit.*; T. WISCHMEYER, *Artificial Intelligence and Transparency cit.*; T. SPEITH, *A review of taxonomies of expleinable artificial intelligence (XAI) cit.*; C. MESKE, E. BUNDE, J. SCHNEIDER, M. GERSCH, *Explainable artificial intelligence cit.*

[626] Cfr. P.P. ANGELOV, E.A. SOARES, R. JIANG, N.I. ARNOLD, P.M. ATKINSON, *Explainable artificial intelligence: an analytical review*, in *WIREs Data Mining & Knowledge Discovery*, 11, 5, 2021, https://bit.ly/3e9O2DE (2 settembre 2022); V. BELLE, I. PAPANTONIS, *Principles and practice of explainable machine learning*, 2020, https://arxiv.org/abs/2009.11698 (2 agosto 2022); M. LANGER, D. OSTER, T. SPEITH, H. HERMANNS, L. KÄSTNER, E. SCHMIDT ET AL., *What do we want from Explainable Artificial Intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research*, in *Artificial Intelligence*, 7, 2021, https://doi.org/10.1016/j.artint.2021.103473 (2 agosto 2022); J.A. MCDERMID, Y. JIA, Z. PORTER, I. HABLI, *Artificial intelligence explainability: the technical and ethical dimensions*, in *Philosophical Transaction Royal Society A*, 2021, doi:379(2207):20200363; D. MINH, X. WANG, Y.F. LI T.N. NGUYEN, *Explainable artificial intelligence: a comprehensive review*, in *Artificial Intelligence Review*, 55, 5, 2022, p. 3503-3568; G. VILONE; L. LONGO,

nessuna di queste categorizzazioni può dirsi completa e coerente, secondo uno studio presentato nel giugno 2022, che significativamente si poneva l'obiettivo di passare in rassegna le tassonomie proposte fino a quel momento[627]. La ricostruzione dei principali avanzamenti nel campoqui offerta, allora, è quindi svolta senz'animo di completezza – un obiettivo che, del resto, pare difficilmente raggiungibile - e al solo fine di fornire un primo quadro introduttivo e generale.

Come già detto al paragrafo precedente, la varietà delle iniziative con cui si tenta di limitare l'opacità di alcune tecnologie intelligentiriflette la molteplicità dei possibili significati attribuibili al concetto di spiegazione, e, dunque, degli obiettivi che ciascun approccio all'*explainable artificial intelligence*, nel concreto, si pone.Da questo punto di vista, una sintesi estrema tra le varie tassonomie proposte permette di identificare tre criteri principali in base ai quali suddividere le metodologie adottate più di frequente per incrementare la comprensibilità dell'intelligenza artificiale. In primo luogo, può distinguersi tra approcci *model-agnostic* e *model-specific*: i primi volti allo sviluppo di tecniche per la spiegazione degli *output* di tutti o di una pluralità di sistemi, i secondi destinati fin dall'origine a migliorare la comprensibilità dei risultati di un sistema determinato[628]. In secondo luogo, è comune differenziare tra*global-methods*, che assumono l'obiettivo di spiegare l'intero funzionamento di una determinata applicazione dell'intelligenza artificiale, e *local-methods*, diretti a ricavare una spiegazione specifica del singolo *output* del sistema cui sono applicati[629]. In terzo luogo, si individuano metodi che si basano sullo sviluppo di un surrogato – tecnologico, umano o meramente teorico– del sistema di cui ci si propone di migliorare la comprensibilità, al quale viene fatto svolgere lo stesso compito, al fine di confrontare i due esiti (da taluni indicati, appunto, come *surrogate methods*)[630], e strategie che mirano a spiegare una tecnologia *black-box* senza la costruzione di percorsi alternativi, agendo sulle modalità con cui questa comunica all'esterno i propri risultati: particolarmente interessanti, come si dirà, sono alcune strategie di *data visualization* applicate alle reti neurali (tanto che questo insieme di tecniche è stato indicato con l'espressione *visualization methods*)[631]. È agevole notare come ciascuna di queste categorie implichi l'adesione tendenziale a uno dei possibili significati del concetto di spiegazione già visti: l'idea di un'*explainable artificial intelligence model-agnostic*, ad esempio, si sposa più facilmente con l'obiettivo di una motivazione dell'*output* di un sistema elaborata *ex post*,

---

Classification of Explainable Artificial Intelligence Methods through Their Output Formats, in *Machine Learning and Knowledge Extraction*, 3, 3, 2021, p. 615-661.

[627] T. SPEITH, *A review of taxonomies of explainable artificial intelligence (XAI) cit.*, peraltro a cura di uno degli autori di una di tali proposte.

[628] Cfr. in particolare V. BELLE, I. PAPANTONIS, *Principles and practice of explainable machine learning cit.*

[629] Cfr. ad esempio G. VILONE, E. LONGO, *Explainable artificial intelligence: a systematic review*, p. 15 ss.

[630] Utilizzano l'espressione, tra i molti, P.P. ANGELOV, E.A. SOARES, R. JIANG, N.I. ARNOLD, P.M. ATKINSON, *Explainable artificial intelligence: an analytical review cit.*, p. 7 ss.

[631] Cfr. in particolare G. ALICIOGLU, B. SUN, *A survey of visual analytics for Explainable Artificial Intelligence methods*, in *Computers & Graphics*, 102, 2022, p. 502-520.

indipendente dalle caratteristiche della tecnologia che vi ha condotto; gli approcci che mirano a ottenere una conoscenza globale del funzionamento del sistema, invece, assumono l'interpretabilità della singola tecnologia coinvolta come punto di partenza per la costruzione di una spiegazione. Si tratta di ragionamenti che è possibile riprodurre pressoché per ogni innovazione o sperimentazione tecnologica appartenente al filone della XAI.

Definite queste categorie generali – il cui valore, in ogni caso, rimane meramente orientativo, vista la commistione tra concezioni differenti che caratterizza il campo dell'*explainable artificial intelligence*, al pari di ogni altro filone dell'intelligenza artificiale[632] – pare doveroso passare brevemente in rassegna gli spunti più significativi offerti, negli ultimi anni, da ricerca e industria nel campo. Tra le strategie più rilevanti per l'elaborazione di spiegazioni delle tecnologie *black-box* possono menzionarsi, senz'animo di completezza:

- Metodi per ottenere delle informazioni in formato numerico con cui dare all'*output* un significato[633]. Sono state avanzate diverse proposte basate su forme di *input perturbation*, in cui il valore assegnato dal sistema a determinate variabili dell'*input*è consapevolmente manipolato, e il risultato ottenuto utilizzato per comprendere quali variabili abbiano avuto un'influenza maggiore nella formulazione dell'*output* originario[634]. Altre tecniche si basano sullo sviluppo di sistemi più semplici, considerati spiegabili, che svolgono le stesse funzioni di quello originale, al fine di ottenerne una comprensibile approssimazione del comportamento. L'intervallo numerico tra i due *output* è poi utilizzato al fine di valutare la correttezza di quello elaborato dal sistema *black-box*[635].
- Metodi che estraggono spiegazioni c.d. *rule-based* (come alberi decisionali, ragionamenti sillogistici, o percorsi decisionali espressi in notazione logica) dal funzionamento del sistema. La base di partenza, generalmente, è rappresentata da un dataset di decisioni o

[632] Un esempio particolarmente noto di come, nella pratica, le strategie di XAI coinvolgano diverse delle tecnologie e degli obiettivi con cui si tenta di sistematizzarle è rappresentato da LIME, sistema che si propone di spiegare ogni algoritmo di classificazione elaborando un modello interpretabile e spiegabile della singola valutazione fornita da quest'ultimo e presentando, spesso, le informazioni così ricavate sul classificatore con tecniche di *data visualization*, cfr. M.T. RIBEIRO, S. SINGH, C. GUESTRIN, *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*, 2016, http://arxiv.org/abs/1602.04938 (4 agosto 2022).

[633] Cfr. ad es. G. VILONE, L. LONGO, *Classification of Explainable Artificial Intelligence Methods through Their Output Formats cit.*, p. 620 ss.

[634] Cfr. ad es. M. ROBNIK-SIKONJA, I. KONONENKO, *Explaining Classifications ForIndividual Instances*, in *IEEE Transactions on Knowledge and Data Engineering*, 20, 5, 2008, p. 589-600; P. CORTEZ, M.J. EMBRECHTS, *Opening black box data mining models using sensitivity analysis*, in *Proceedings of the Symposium on Computational Intelligence and Data Mining (CIDM)*, Paris, France, 11–15 aprile 2011, p. 341–348.

[635] Si vedano ad es. S. TAN, R. CARUANA, G. HOOKER, Y. LOU, *Distill-and-Compare: Auditing Black-Box Models Using Transparent Model Distillation*, in A.A.V.V., *Proceedings of the Conference on AI, Ethics, and Society*, New Orleans, 2018, p. 303–310; S.M. LUNDBERG, S.I. LEE, *A unified approach to interpreting model predictions*, in *Advances in Neural Information Processing Systems - Neural Information Processing Systems Foundatio*n, Long Beach 4–9 December 2017, p. 4765–4774.

predizioni prese dal sistema e dei relativi *input*, analizzato con strumenti di intelligenza artificiale al fine di elaborare regole in grado di rappresentare una spiegazione nel senso appena detto, la cui bontà è verificata in base alla loro capacità di rappresentare, in tutto o in parte, gli *output* raccolti nel *dataset* di riferimento[636].

- Metodi che offrono spiegazioni *controfattuali*, utilizzando le tecniche, già menzionate, che ambiscono a capire quali, tra i dati di *input*, siano i più significativi per il risultato fornito al sistema. La spiegazione consiste nella risposta a una *what-if question*, ovvero in informazioni su quali caratteristiche differenti da quelle di partenza sarebbero state decisive per portare il sistema a un *output* diverso[637].
- Metodi in grado di fornire spiegazioni in formato grafico, basati su strategie di *data visualization*[638]. Particolarmente significative paiono alcune tecniche utilizzate a supporto di sistemi di riconoscimento e categorizzazione di immagini basati sul *deep learning*, in grado di evidenziare le aree dell'immagine utilizzata come *input* che più hanno influito sulla decisione finale[639]. Altre applicazioni, invece, legate più strettamente a una concezione di spiegazione come interpretabilità del funzionamento interno del sistema, producono rappresentazioni in formato grafico delle aree della rete maggiormente coinvolte nell'elaborazione dell'*input*[640].
- Metodi in grado di fornire spiegazioni in formato testuale, combinando tecnologie di elaborazione del linguaggio naturale a tecniche di *explainable artificial intelligence* vera e propria come quelle viste finora, il cui sviluppo potrebbe rendere più accessibile anche a utenti non esperti la spiegazione di volta in volta elaborata, a prescindere dalla sua natura[641].

[636]H. BRIDE, J. DONG, J. S. DONG, Z. HÓU, *Towards dependable and explainable machine learning using automated reasoning*, in *Proceedings of the International Conference on Formal Engineering Methods*, Gold Coast, Australia, 12–16 November 2018, p. 412–416; S. KRISHNAN, E. WU, *PALM: Machine Learning Explanations For Iterative Debugging*, in *Proceedings of the 2nd workshop on Human-In-the-Loop data analytics,* in *ACM, Chicago IL USA*, 2017, p. 1-6, doi:10.1145/3077257.3077271.

[637] Cfr. in particolare Y. CHOU, C. MOREIRA, P. BRUZA, C. OUYANG, J. JORGE, *Counterfactuals and causability in explainable artificial intelligence: theory, algorithms, and applications*, in *Information Fusion*, 81, 2022, p. 59-83.

[638] Per una panoramica delle principali innovazioni riconducibili a questa famiglia cfr. ancora G. ALICIOGLU, B. SUN, *A survey of visual analytics for Explainable Artificial Intelligence cit.*

[639] Cfr. Y.G. CHAN, E. BERTINI, L.G. NONATO, B. BARR, C.T. SILVA, *Melody: Generating and Visualizing Machine Learning Model Summary to Understand Data and Classifiers Together*, http://arxiv.org/abs/2007.10614 (4 agosto 2022).

[640]M. KAHNG, P.Y. ANDREWS, A. KALRO, D.H. CHAU, *ActiVis: Visual Exploration of Industry-Scale Deep Neural Network Models*, in *IEEE Transactions on Visualization and Computer Graphics*, 24, 1, 2018, p. 88-97.

[641]S. BARRATT, *Internet: Neural introspection for interpretable deep learning, in Proceedings of the Symposium on Interpretable Machine Learning*, Long Beach, CA, USA, 7 December 2017, p. 47–53; I. GARCÍA-MAGARIÑO, R. MUTTUKRISHNAN, J. LLORET, *Human-Centric AI for Trustworthy IoT Systems With Explainable Multilayer Perceptrons*, in *Access*, 7, 2019, p. 125562–125574.

Allo stato dell'arte, come già detto, gli avanzamenti dell'*explainable artificial intelligence* avvengono prevalentemente nel mondo della ricerca, a causa della novità del settore. Le applicazioni pratiche, pur già molto significative, paiono, invece, agli inizi[642]. L'interesse del mondo produttivo e dei regolatori pubblici per il campo non sembra potersi mettere in discussione: ne dà una dimostrazione efficace il numero di analisi, *position paper*, e dichiarazioni d'intenti sul tema della spiegabilità o trasparenza dell'intelligenza artificiale pubblicati da società multinazionali, società di consulenza e gruppi di esperti appositamente costituiti, spesso di nomina pubblica[643]. L'opacità di alcune tecnologie basate sull'intelligenza artificiale è percepita come uno dei principali problemi da risolvere per poter godere appieno dei vantaggi che tali innovazioni rappresentano. I potenziali effetti positivi dello sviluppo del campo dell'*explainable artificial intelligence* sono intuitivi, e sono stati evidenziati da più parti: l'utilizzo delle tecnologie intelligenti nei processi decisionali ne risulterebbe facilitato, permettendo all'operatore umano di valutare con maggior completezza l'indicazione della macchina; gli esiti negativi di eventuali *output* errati o discriminatori sarebbero più facilmente individuabili, evitabili ed eventualmente correggibili; il livello di sicurezza, effettiva e percepita, di molte tecnologie avanzate crescerebbe, favorendo il loro impiego in contesti ad alto rischio; l'allocazione di eventuali responsabilità per malfunzionamenti e conseguenze indesiderate risulterebbe più agevole; i sistemi risulterebbero più facilmente migliorabili da parte dei tecnici di riferimento[644].

A conclusione di questo breve inquadramento iniziale, preme evidenziare che, in ogni caso, lo stato dell'arte tecnico-scientifico in materia di *explainable artificial intelligence* non consente di considerare totalmente risolte le questioni sollevate dall'opacità di talune applicazioni dell'intelligenza artificiale.Infatti, se è doveroso sottolineare come siano già stati conseguiti risultati importanti, e come i rapidi progessi nel campo rendano legittimo sperare nel raggiungimento di

[642] La varietà di applicazioni già in uso non è, comunque, di certo da sottovalutare, cfr. I. AHMED, G. JEON, F. PICCIALLI, *From artificial intelligence to explainable artificial intelligence in Industry 4.0: a survey on what, how, and where*, in *IEEE Transactions on Industrial Informatics*, 18, 8, p. 5031-5042, 2022, doi:10.1109/TII.2022.3146552 (4 agosto 2022).

[643] Il tema, ad esempio, è trattato dalle più volte menzionate EC HIGH LEVEL EXPERT GROUP ON AI, *Ethics Guidelines for Trustworthy Artificial Intelligence cit.*, p. 25. Si vedano anche, *ex multis*, Deloitte, *Bringing transparency and ethics in AI*, 2019, https://bit.ly/3Qek1QE (4 agosto 2022); Google, *AI Explainability Whitepaper*, 2020, https://bit.ly/3CTput7 (4 agosto 2022).

[644] Evidenziano questi elementi, in vario modo, A. HABAYEB, *Explainable AI isn't enough; we need understandableAI*, in *Techopedia.com*, https://bit.ly/3CWvKA8 (4 agosto 2022); O. G. TALCIN, *5 significant reasons why explainable AI i san existential needs for humanity*, in *Towards Data Science*, 28 dicembre 2020 (4 agosto 2020); F. PASQUALE, *The Black-Box Society: The Secret Algorithms That Control Money and Information cit.;* S. QUINTARELLI, F. COREA, F. FOSSA, A. LOREGGIA, S. SAPIENZA, *AI: Profili etici. Una prospettiva etica sull'intelligenza artificiale: principi, diritti e raccomandazioni cit.*, le appena citate HIGH LEVEL EXPERT GROUP ON AI, *Ethics Guidelines for Trustworthy Artificial Intelligence cit.* e Deloitte, *Bringing transparency and ethics in AI cit.*; Google, *AI Explainability Whitepaper cit.* Per un interessante studio sul legame tra finalità morali e sociali considerate prevalenti e la scelta di determinati approcci di XAI, cfr. J.A. MCDERMID, Y. JIA, Z. PORTER, I. HABLI, *Artificial intelligence explainability: the technical and ethical dimensions cit.*

ulteriori innovazioni d'impatto nel breve e medio periodo, pare altrettanto necessario, a questo punto, rimarcare i limiti del settore, almeno in questo momento storico.

In primo luogo, è doveroso sottolineare che esiste un netto *trade-off* tra la spiegabilità dei sistemi e la loro efficacia in diversi settori applicativi. Le tecnologie *black-box* più complesse, *in primis* quelle riconducibili alla famiglia delle reti neurali, garantiscono prestazioni irraggiungibili per i sistemi *model-based*, soprattutto nei campi in cui l'utilizzo di questi ultimi si è sempre rivelato particolarmente difficile (è il caso, ad esempio, dei già più volte menzionati ambiti del riconoscimento d'immagini o della traduzione automatica)[645]. Dunque, non è ipotizzabile sostituire, integralmente e in tutti i settori, gli algoritmi caratterizzati dall'opacità con altri maggiormente trasparenti, se non al costo di una diminuzione della qualità delle prestazioni spesso radicale. Si tratta, del resto, della principale ragione dell'attuale picco d'interesse verso l'*explainable artificial intelligence*, chesi pone, appunto, l'obiettivo di sviluppare sistemi maggiormente trasparenti senza diminuirne al contempo l'efficacia, o di rinvenire strategie per rendere maggiormente comprensibili le tecnologie *black-box* senza alterarne struttura e funzionamento.

In secondo luogo, deve rilevarsi come le spiegazioni offerte dai visti metodi per l'ottenimento di un'*explainable artificial intelligence* sembrino presentare, in alcuni casi, diverse debolezze[646]. Ad esempio, studi recenti hanno messo in luce come alcune delle strategie utilizzate per identificare quali, tra i dati di *input*, abbiano un peso più rilevante per la determinazione dell'*output* risultino sensibili a variazioni minime, che non appaiono dotate di senso all'operatore umano. La modifica di alcuni *pixel* di un'immagine può portare a variazioni del risultato cui giunge un sistema di riconoscimento e classificazione e, allo stesso modo, può influenzare la spiegazionefornita con tecniche di *explainable artificial intelligence*[647]. La ricerca in materia è ancora in una fase precoce, e la complessità del tema impone cautela nell'estrarre considerazioni generali da studi che, generalmente, riguardano una specifica applicazione o una famiglia di tecnologie piuttosto ristretta.

---

[645] Si rimanda, tra gli studi a riguardo, ai già citati R.V. YAMPOLSKIY, *Unexplainability and incomprensibility cit.*; A. ADADI, M.BERRADA, *Peeking Inside the Black-Box cit.*; S. SARKAR ET AL., *Accuracy and Interpretability Trade-offs in Machine Learning Applied to Safer Gambling*, in *CEUR Worskshop Proceedings*, 2016, p. 1773 ss.

[646] S. KRISHNA, T. HAN, A. GU, J. POMBRA, S. JABBARI, S. WU, *The Disagreement Problem in Explainable Machine Learning: A Practitioner's Perspective*, 2022, https://arxiv.org/abs/2202.01602 (4 agosto 2022); W. HRYNIEWSKA, P. BOMBIŃSKI, P. SZATKOWSKI ET AL., *Do not repeat these mistakes -- a critical appraisal of applications of explainable artificial intelligence for image based COVID-19 detection*, 2020, https://europepmc.org/article/PPR/PPR274054 (arXIv preprint); Z. BUÇINCA, P. LIN, K.Z. GAJOS, E.L. GLASSMAN, *Proxy tasks and subjective measures can be misleading in evaluating explainable AI systems*, in *Proceedings of the 25th International Conference on Intelligent User Interfaces - ACM, Cagliari* (Italy), 2020, p. 454-464, https://dl.acm.org/doi/10.1145/3377325.3377498 (4 agosto 2022); H. DE BRUIJN M. WARNIER, M. JANSSEN, *The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making*, in *Government Information Quarterly*, 39, 2, 2022, 101666.

[647] Sul tema della sensibilità dei sistemi di riconoscimento delle immagini a variazioni impercettibili per l'essere umano cfr. in particolare A. ROSENFELD, R. ZEMEL, J.K. TSOTSOS, *The Elephant in the Room*, 2018, https://arxiv.org/abs/1808.03305 (4 agosto 2022); una panoramica dei principali studi che, direttamente o meno, affrontano il tema della sensibilità delle tecniche di *explainable artificial intelligence* alla modifica dei dati di input può rinvenirsi in G. VILONE, E. LONGO, *Explainable artificial intelligence: a systematic review cit.*, p. 44 ss.

Pare potersi affermare, però, che, talvolta, alcune caratteristicheproblematiche di certe tecnologie *black-box*, che generano la possibilità, seppur ridotta, di errori difficilmente identificabili e non comprensibili per l'essere umano, si ripercuotano anche sulle tecniche utilizzate per estrarre una spiegazione. La quale, dunque, sembra esposta allo stesso margine d'errore difficilmente razionalizzabile. È stato evidenziato, inoltre, che le spiegazioni fornite dai diversi metodi di *explainable artificial intelligence* possono essere utilizzate per ricostruire e plagiare il sistema o per attaccarne il funzionamento[648]. La circostanza, del resto, non ècerto un tratto caratteristico dell'intelligenza artificiale: a un maggior disvelamento di informazioni riguardanti una determinata applicazione tecnologica si è quasi sempre accompagnato un incremento dei rischi dal punto di vista della sicurezza informatica, fin dal principio della rivoluzione digitale[649].

In terzo luogo, preme ribadire che le informazioni che l'*explainable artificial intelligence* permette di ricavare corrispondono a una vasta gamma di possibili varianti del concetto di spiegazione, alcune delle quali potrebbero apparire, a taluni osservatori, estremamente parziali, meramente indiziarie, prive di connessione con la tecnologia di partenza e non supportate da elementi sufficienti a riprova della loro bontà. Più in generale, ogni spiegazione, anche elaborata da esseri umani, non è che un modello che, nell'intento di rendere comprensibile e comunicabile la realtà di volta in volta spiegata, ne fornisce una rappresentazione inevitabilmente semplificata, imperfetta ed esposta a un margine d'errore[650]. Come già accennato, ogni ragionamento sulla spiegabilità dei risultati di un sistema intelligente è prima di tutto un ragionamento sul tipo di spiegazione che si considera sufficiente, in generale o in un determinato contesto applicativo. Come si vedrà nei prossimi paragrafi, questa constatazione è valida, in primo luogo, per il ragionamento giuridico in materia di regolazione dell'intelligenza artificiale.

3.3 È sempre necessario che l'intelligenza artificiale sia trasparente? Alcune considerazioni su spiegabilità e bilanciamento con altri diritti e interessi

L'analisi, fino ad ora, ha dato per presupposto che una spiegazione sia necessaria, e percepita come tale in primo luogo dagli utenti di sistemi *black-box*. Giunti a questo punto, è opportuno dar conto dell'esistenza di risultanze scientifiche in materia in parte discordanti, in particolare quando gli interessi in gioco sembrino d'importanza limitata. Un'indagine condotta, già nel 2012, sulle sensazioni suscitate dall'interazione con sistemi *black-box* in un ampio gruppo di persone di diversa

[648] S. MILLI, L. SCHMIDT, A. DRAGAN, M. HARDT, *Model Reconstruction from Model Explanations*, http://arxiv.org/abs/1807.05185 (4 agosto 2022); R. SHOKRI, M. STROBEL, Y. ZICK, *On the Privacy Risks of Model Explanations*, 2019, https://arxiv.org/abs/1907.00164 (4 agosto 2022).

[649] A. BURT, *The AI transparency paradox*, in *Harvard Business Review*, 13 dicembre 2019, https://hbr.org/2019/12/the-ai-transparency-paradox (4 agosto 2022).

[650] Cfr. ancora T. MILLER, *Explanation in artificial intelligence cit.*; J. C. PITT (A CURA DI), *Theories of explanation cit.*; W. G. LYCAN, *Explanation and epistemology cit.*

estrazione sociale e livello culturale, ad esempio, ha messo in luce che, se il sistema è applicato in un ambito di rilievo circoscritto, il bisogno di una spiegazione dei suoi risultati è avvertito solo da una minoranza degli utenti[651]. In particolare, la ricerca prendeva in considerazione gli algoritmi utilizzati da diverse famose piattaforme di *video-sharing* ed*e-commerce* per la raccomandazione di contenuti agli utenti, spesso consistenti in complessi sistemi di apprendimento automatico. Interrogati sul loro bisogno di conoscere la logica in base a cui tali raccomandazioni venivano formulate, parte degli utenti ha fornito risposte come «I don't really care how things work, just that they do» o «It wouldn't really bother me if I didn't know, as long as it works»[652]. Alcuni degli argomenti utilizzati per giustificare tale mancanza d'interesse riguardavano l'importanza limitata del contesto di riferimento o la semplice abitudine a tali sistemi, con risposte come: «I think it's mostly because I don't really feel like it affects my life that much» o «I've been using these things so long that I just generally tend to trust them, or not»[653]. Sulla stessa linea, altri utenti hanno dichiarato di reputare lo sforzo necessario per ottenere tale spiegazione eccessivo rispetto al limitato vantaggio che essa avrebbe apportato alle loro esistenze[654]. Per quanto gli studi sperimentali di questo genere non siano particolarmente numerosi[655], una conferma della tendenziale bontà di questi risultati può ricavarsi, ormai, dall'esperienza quotidiana di ciascuno, posta la diffusione capillare di questo genere di applicazioni e la disinvoltura con cui quasi tutti le utilizzano, spessocon un'idea estremamente approssimativa del loro funzionamento. Chiaramente, nel valutare l'ipotesi di costruire presidi giuridici per garantire una soglia minima di spiegabilità dell'intelligenza artificiale questi dati non possono essere ignorati. Non avrebbe senso, infatti, porre ostacoli normativi allo sviluppo e all'applicazione di determinate applicazioni tecnologiche a tutela di prerogative che sembrano oggettivamente di rilevanza limitata e non appaiono meritevoli di protezione ai loro stessi titolari.Tuttavia, l'argomento pare spendibile solo con cautela, poiché non sempre l'individuo è in grado di valutare il potenziale lesivo di determinate azioni e situazioni (l'ignoranza pressoché totale, fino alla seconda metà del Novecento, sui danni da fumo di sigaretta non ha certo messo al riparo da essi generazioni di fumatori). Né è sempre agevole individuare, sul piano oggettivo, quali ambiti o applicazioni tecnologiche abbiano effettivamente un'importanza

[651] A. BUNT, M.LOUNT, C. LAUZON, *Are explanations always important? A study of deployed, low-cost intelligent interactive systems*, in *Proceedings of the 2012 ACM international conference on IUI*, 2012, p. 169 ss.
[652] A. BUNT, M.LOUNT, C. LAUZON, *Are explanations always important?*, p. 173.
[653] *Ibidem*, p. 176.
[654] Due degli intervistati, ad esempio, hanno dichiarato: «I just don't HAVE to know it» e «Are there options for me after I know this information? Like, I know how it works, so I can adjust them, so then it works better for me. Without the second step, it doesn't make any sense» cfr. *Ibidem*, p. 174.
[655] Possono menzionarsi, ad esempio, M.T. DZINDOLET, S.A. PETERSON, R.A. POMRANKY, L.G. PIERCE, H.P. BECK, *The role of trust in automation reliance*, in *International Journal of Human-Computer Studies*, 58, 6, 2003, p. 697–718; B.Y. LIM, A.K. DEY, *Assessing demand for intelligibility in context-aware applications*, in *Proceedings of the 11th international conference on ubiquitous computing*, Orlando, 2009, p. 195–204, peraltro piuttosto risalenti.

limitata. Sistemi di raccomandazione analoghi a quelli utilizzati dalle piattaforme, ad esempio, sono stati tra i principali responsabili della crescente polarizzazione politica sulle reti sociali, sembrata in grado di minare lo stesso regolare svolgimento di alcune elezioni[656]. La reale portata degli interessi in gioco, dunque, dovrà di certo avere un peso, come si dirà, nella regolazione dell'intelligenza artificiale, ma non può essere un facile argomento con cui accantonare o sminuire le questioni connesse all'opacità di determinate tecnologie.

La necessità e utilità di una spiegazione dell'intelligenza artificiale sembrano potersi mettere in discussione anche da un punto di vista diametralmente opposto: quando l'intelligenza artificiale sia utilizzata in contesti in cui sono coinvolti valori primari, come la tutela della vita e della salute. È lecito chiedersi, infatti, se abbia senso mettere in discussione in tali contesti l'utilizzo di sistemi *black-box* per i quali anche le spiegazioni ottenibili con le tecniche di *explainable artificial intelligence* non sembrino sufficienti a taluni osservatori, ma la cui efficacia, misurabile su base statistica, risulti ineguagliata da tecnologie d'altro genere[657]. Il tema sarà analizzato più nel dettaglio negli ultimi capitoli del lavoro, dedicati all'analisi di alcune applicazioni dell'intelligenza artificiale in ambito sanitario. Dal punto di vista delle ipotesi di regolazione delle tecnologie avanzate, in ogni caso, la risposta non può che essere analoga a quella vista poche righe sopra: l'importanza dei valori concretamente in gioco deve di certo essere tenuta in considerazione, e nessuna disciplina giuridica può avere il risultato di ostacolare la tutela di beni primari come la vita o l'integrità fisica. D'altro canto, la delicatezza del campo di applicazione di alcuni strumenti basati sull'intelligenza artificiale non può rappresentare un impedimento allo sviluppo di una loro regolazione che tuteli la centralità dell'essere umano e dei diritti fondamentali, anche dal punto di vista dell'eventuale introduzione di

[656] Si rimanda, *ex multis*, ai già citati C. SUNSTEIN, *#Republic.com: divided democracy in the age of social media cit.*; J. M. BALKIN, *Free speech in the algorithmic society cit.*; G. PITRUZZELLA, O. POLLICINO, S. QUINTARELLI, *Parole e potere. Libertàd'espressione, hate speech e fake news cit.*; E. PARISIER, *The filter bubble cit.*; R. GORWA, R. BINNS, C. KATZENBACH, *Algorithmic content moderation: Technical and political challenges in the automation of platform governance cit.*; E. LLANSÒ, J. VON HOBOKEN, P. LEERSSEN, J. HARAMBAM, *Artificial intelligence, content moderation and freedom of expression cit.*; CAMBRIDGE CONSULTANTS, *Report produced on behalf of Ofcom - Use of AI in online content moderation cit.*; C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni cit.*, p. 110 ss. e *Costituzione e intelligenza artificiale cit.*, p. 713 ss; M. FASAN, *Intelligenza artificiale e pluralismo: uso delle tecniche di profilazione nello spazio pubblico democratico cit.*, p. 107 ss.; H. MARSHALL, A. DRIESCHOVA, *Post-truth politics in the UK's Brexit referendum cit.*; R. KÜBLER, K. PAUWELS, K. MANKE, *How Social Media Drove the 2016 US Presidential Election cit.*; Y. TSFATI, H.G. BOOMGAARDEN, J. STRÖMBÄCK, R. VLIEGENTHART, A. DAMSTRA, E. LINDGREN, *Causes and consequences of mainstream media dissemination of fake news cit.*; A. R. DOSHI, S. RAGHAVAN, R. WEISS, E. PETITT, *How the supply of fake news affected consumer behavior during the 2016 US election cit*.

[657] Alcuni esempi particolarmente noti riguardano, allo stato dell'arte, determinati utilizzi delle reti neurali per il riconoscimento di immagini cliniche, cfr. A. HEKLER ET AL., *Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images*, in *European Journal of Cancer*, 118, 2019, p. 91 ss.; P. RAJPURKAR ET AL., *CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning*, 2017, arXiv:1711.0522. Riguardo al trade-off ancora presente, allo stato dell'arte, tra efficacia e comprensibilità di determinati algoritmi, cfr. R.V. YAMPOLSKIY, *Unexplainability and incomprensibility cit.*; A. ADADI, M. BERRADA,*Peeking Inside the Black-Box cit.*; S. SARKAR ET AL., *Accuracy and Interpretability Trade-offs in Machine Learning Applied to Safer Gambling*, in *CEUR Worskshop Proceedings*, 2016, p. 1773 ss.

determinati requisiti minimi di spiegabilità. Al contrario, proprio il potenziale rilievo primario dei valori in gioco rende particolarmente urgente la riflessione giuridica sul tema.

3.4 Il contenuto del diritto alla spiegazione e i primi, parziali riconoscimenti nel diritto positivo

Esaurito l'inquadramento generale sui sistemi di intelligenza artificiale *black-box* che ha occupato gli ultimi paragrafi, si passi all'analisi delle soluzioni che il diritto può offrire alle questioni da essi sollevate e, in particolare, alla possibilità di riconoscere, nella spiegabilità dei sistemi, un vero e proprio diritto fondamentale.

Tra i nuovi diritti teorizzati in questo lavoro, il diritto alla spiegazione è di certo quello che è stato, negli ultimi anni, maggiormente discusso dalla letteratura filosofica e giuridica[658]. Si possono ipotizzare due ragioni essenziali di questo picco d'interesse: in primo luogo,la particolare evidenza degli interessi primari coinvolti, posto che il tema della spiegazione – o di una sua eventuale mancanza – richiama immediatamente alcune tra le garanzie giuridiche più consolidate, come il diritto alla motivazione ed eventuale riesame dei provvedimentiamministrativi o giudiziari; in secondo luogo, la sua parziale positivizzazione in alcuni ordinamenti, *in primis* quello dell'Unione Europea. Tali iniziative normative, infatti, hanno stimolato un vivace dibattito dottrinale, per l'innovazione che rappresentavano e per la genericità, secondo alcuni eccessiva, della loro formulazione.

Di un possibile riconoscimento di un *right to an explanation* hanno fatto parlare soprattutto alcune norme del GDPR, in particolare gli artt. 13, 14, 15 e 22 e il Cons. 71, già più volte citati nel corso del lavoro. L'ambito di riferimento di tali norme, com'è noto, sono i trattamenti di dati che consistano in una decisione totalmente automatizzata, inclusa la profilazione, riguardante una persona fisica. In linea di principio, l'art. 22 par. 1 vieta tali trattamenti ogni qualvolta producano «effetti giuridici che la riguardano o incidano in modo analogo significativamente sulla sua

[658] Cfr. ad es. S. WACHTER, B. MITTELSTADT, L. FLORIDI, *Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation*, in *International Data Privacy Law*, 7, 2, 2017, p. 76–99;B. GOODMAN, S. FLAXMAN, *European Union regulations on algorithmic decision-making and a "right to explanation"*, in AI Magazine, 2017, p. 50-57; L. EDWARDS, M. VEALE, *Enslaving the algorithm: from a "right to an explanation" to a "right to better decisions"?*, in *IEEE Security & Privacy*, May-June 2018, p. 46-54; K. VRENDENBURGH, *The right to explanation*, in *The Journal of Political Philosophy*, 30, 2, 2022, p. 209-229; A.D. SELBST, J. POWLES, *Meaningful information and the right to explanation*, in *International data privacy law*, 7, 4, 2017; M. E. KAMINSKI, *The right to explanation, explained*, in *Berkeley Technology Law Journal,* 34, 1, 2019, p. 189 ss.; B. CASEY, A. FARHANGI, R. VOGL, *Rethinking explainable machines: the GDPR's "right to explanation" debate and the rise of algorithmic audits in enterprise*, in *Berkeley Technology Law Journal,* 34, 1, 2019, p. 143 ss.; J. GACUTAN, N. SELVADURAI, *A statutory right to explanation for decisions generated using artificial intelligence*, in *International Journal of Law and Information Technology*, 28, 2020, p. 193-216; G. MALGIERI, *Automated decision-making in the EU Member States: the right to explanation and other "suitable safeguards" in the national legislations*, in *Computer law & security review*, 25, 2019.C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni cit.* e *Costituzione e intelligenza artificiale cit.*; A. SIMONCINI, *L'algoritmo incostituzionale cit*.

persona»[659]. Il secondo paragrafo della norma pone alcune eccezioni al divieto, stabilendo che tal genere di trattamenti sia permesso quando necessario per la conclusione o l'esecuzione di un contratto tra titolare e interessato, si basi sul consenso di quest'ultimo o sia esplicitamente permesso dal diritto dell'Unione o di uno stato membro[660]. L'ultimo paragrafo pone ulteriori vincoli a queste eccezioni qualora il trattamento abbia ad oggetto dati c.d. particolari, per i quali sono previsti presidi specifici in ragione della loro intrinseca delicatezza[661]. In tali casi, le uniche basi giuridiche possibili per il trattamento sono il consenso esplicito dell'interessato (art. 9 lett. a) o il perseguimento di motivi di rilevante interesse pubblico sulla base del diritto dell'Unione o degli stati membri (art. 9 lett. g). L'art. 22, inoltre, quando la decisione automatizzata è permessa impone sempre l'adozione di misure adeguate per la «tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato», precisando, per i casi in cui il trattamento sia necessario alla conclusione o esecuzione di un contratto, o si basi sul consenso dell'interessato, che esse devono consistere almeno «nel diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione». Gli artt. 13 par. 1 lett. f), 14 par. 1 lett. g) e 15 par. 1 lett. h), relativi alle informazioni da fornire all'interessato del trattamento, richiamano l'art. 22, stabilendo che l'interessato deve sempre essere informato dell'esistenza del trattamento automatizzato e, in tali casi, devono essergli fornite «informazioni significative sulla logica utilizzata»[662]. Il quadro ècompletato dal Considerando 71, che, come esempi delle garanzie che è

[659]Si riporta nuovamente (cfr. p. 140, n. 543), per chiarezza, il testo completo dell'art. 22: «1. L'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona. 2. Il paragrafo 1 non si applica nel caso in cui la decisione: a) sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento; b) sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato; c) si basi sul consenso esplicito dell'interessato. 3. Nei casi di cui al paragrafo 2, lettere a) e c), il titolare del trattamento attua misure appropriate per tutelare i diritti, le libertà e i legittimi interessi dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione. 4. Le decisioni di cui al paragrafo 2 non si basano sulle categorie particolari di dati personali di cui all'articolo 9, paragrafo 1, a meno che non sia d'applicazione l'articolo 9, paragrafo 2, lettere a) o g), e non siano in vigore misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato». Per dei commenti cfr. ancora A. ODDENINO, *Decisioni algoritmiche e prospettive internazionali cit.*; A. CAIA, *Art. 22 cit*.

[660] L'ampiezza di tali eccezioni, come già approfondito, è stata estensivamente criticata, cfr. ad es. C. CASONATO, *Costituzione e intelligenza artificiale cit.*, p. 723-724; A. SIMONCINI, *L'algoritmo incostituzionale cit.*, p. 79 ss.

[661] I dati c.d. particolari, in larga misura corrispondenti a quelli definiti, nel regime precedente all'emanazione del GDPR, "dati sensibili", sono le categorie di dati per cui il Regolamento dispone forme rafforzate di tutela, in ragione della loro particolare delicatezza. A delimitarne il perimetro è l'art. 9 par. 1 (rubricato *categorie particolari di dati personali*) che ne vieta, in generale, il trattamento: «È vietato trattare dati personali che rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, o l'appartenenza sindacale, nonché trattare dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona». Il paragrafo successivo elenca le eccezioni al divieto che si risolvono, in sostanza, nelle basi giuridiche del trattamento.

[662] Nello specifico, l'art. 13 del Regolamento individua le informazioni da fornire qualora i dati personali siano raccolti presso l'interessato, l'art. 14 qualora i dati non siano raccolti presso di esso, l'art. 15 qualora sia l'interessato ad esercitare il diritto d'accesso ai dati trattati. Come già riportato, le tre norme, con identica formulazione, impongono di comunicare all'interessato «l'esistenza di un processo decisionale automatizzato, compresa la profilazione di cui all'articolo 22, paragrafi 1 e 4, e, almeno in tali casi, informazioni significative sulla logica utilizzata, nonché

opportuno circondino il trattamento nei casi in cui la decisione automatizzata risulti lecita ai sensi dell'art. 22, indica: «la specifica informazione all'interessato e il diritto di ottenere l'intervento umano, di esprimere la propria opinione, di ottenere una spiegazione della decisione conseguita dopo tale valutazione e di contestare la decisione»[663].

La sola norma a menzionare esplicitamente un diritto alla spiegazione, dunque, è il Considerando 71, disposizione che, però, com'è noto non ha pieno valore precettivo, avendo la funzione di guidare l'interpretazione degli articoli del Regolamento. Il resto delle disposizioni prese in esame, com'è agevole notare, ha una formulazione estremamente generale e, dunque, potenzialmente aperta a una grande varietà di interpretazioni. Queste circostanze hanno portato i primi commentatori a dividersi tra chi sottolineava, in varia misura, l'importanza delle innovazioni introdotte dal GDPR[664] e chi, invece, riteneva eccessiva la cautela dimostrata dal Regolamento verso l'ipotesi di un diritto alla spiegazione, considerando la formulazione adottata nell'articolato eccessivamente timida per rappresentare un'efficace tutela individuale[665]. Le critiche più comuni riguardavano, in primo luogo, l'ambito di applicazione delle norme appena viste, ritenuto troppo limitato, essendo circoscritto ai trattamenti consistenti in decisioni totalmente automatizzate, con l'esclusione, così, di una vastissima gamma di sistemi di supporto alle decisioni degli esseri umani di enorme rilievo pratico, e spesso decisivi per la formulazione di queste ultime[666]. In secondo luogo, le principali voci scettiche sulla portata del diritto alla spiegazione nel GDPR, o sulla sua stessa esistenza giuridica, sottolineavano la vaghezza delle disposizioni del Regolamento, soprattutto riguardo all'espressione «informazioni significative sulla logica utilizzata», riconducibile al concetto di spiegazione solo con un'intensa attività interpretativa[667]. In terzo luogo, vi era chi evidenziava come tale formulazione, anche interpretata estensivamente, imponesse unicamente di fornire informazioni sulle tecnologie coinvolte nel trattamento prima che questo avvenisse, al momento di adempiere a tutti i doveri informativi imposti dal Regolamento. Nulla, invece, permetteva di ricavarne il diritto per l'interessato a ricevere una spiegazione della specifica

---

l'importanza e le conseguenze previste di tale trattamento per l'interessato». Per un commento cfr. di nuovo L. GRIECO, *Informazioni e accesso ai dati personali – artt. 13, 14, 15 cit.*

[663] Per il testo completo del Considerando 71 cfr. *supra*, p. 160, n. 540.

[664] Pur con specificazioni di vario genere, cfr. ad es. le opinioni di G. MAUGERI, G. COMANDÉ, *Why a right to legibility of automated decision making exists in the General Data Protection Regulation*, in *International Data Privacy Law*, 7, 4, 2017, p. 243-265; B. GOODMAN, S. FLAXMAN, *European Union regulations on algorithmic decision-making and a "right to explanation" cit.*

[665] È il caso, in particolare, del noto articolo di S. WACHTER, B. MITTELSTADT, L. FLORIDI, *Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation*, che criticano le scelte lessicali del Regolamento europeo e giudicano insufficiente la formulazione adottata. Si rileva, in ogni caso, che le conclusioni cui gli Autori giungono non sembrano così nette come il titolo del contributo porterebbe a pensare; cfr. in proposito L. EDWARDS, M. VEALE, *Enslaving the algorithm: from a "right to an explanation" to a "right to better decisions"? cit.*

[666] Così ad esempio L. EDWARDS, M. VEALE, *Enslaving the algorithm cit.*,p. 3.

[667] Cfr. in particolare S. WACHTER, B. MITTELSTADT, L. FLORIDI, *Why a right to explanation of automated decision-making does not exis cit.*, p. 83 ss.

decisione concretamente assunta nel caso che lo riguardava, successivamente al trattamento[668]. Infine, si faceva notare, semplicemente, che non esisteva una sola disposizione vincolante del GDPR che menzionasse il diritto alla spiegazione[669].

Coloro che, invece, valorizzavano l'impostazione del Regolamento, sottolineavano come proprio il carattere estremamente generale della formulazione legislativa adottata avrebbe potuto rivelarsi decisivo per permetterne l'applicazione in una pluralità di scenari pratici molto diversi tra loro per il contesto di riferimento, il tipo di tecnologia coinvolta e, di conseguenza, le caratteristiche della spiegazione più adatta[670]. Alcuni commentatori, inoltre, sottolineavano come il GDPR prevedesse strumenti per assicurare l'effettivo rispetto di quanto in esso previsto di inedita intensità e potenziale efficacia, anche dal punto di vista dell'apparato sanzionatorio di cui le autorità di vigilanza degli stati membri potevano eventualmente disporre, e argomentavano che ciò avrebbe rappresentato un importante incentivo per l'effettività dei diritti da esso previsti, anche in materia di spiegazione[671]. La portata delle garanzie del GDPR in materia di decisioni automatizzate è stata valorizzata anche da delle apposite Linee Guida del febbraio 2018 dell'*Article 29 Data Protection Working Party*, poi ratificate dallo *European Data Protection Board*, una volta entrato in vigore il Regolamento[672]. Tali Linee Guida hanno sposato, in generale, un'interpretazione estensiva delle norme del GDPR e, relativamente alle «informazioni significative sulla logica utilizzata» da fornire all'interessato del trattamento, hanno charito che: «The controller should find simple waysto tell the data subject about the rational behind, or the criteria relied on in reaching the decision. The GDPR requires the controller to provide meaningful information about the logic involved, not necessary a complex explanation of the algorithms used or disclosure of the full algorithm. The information provided should, however, be sufficiently comprehensive for the data subject to understand the reasons for the decision»[673].

Preme evidenziare, peraltro, come sullo scenario europeo le norme del GDPR non siano le uniche a garantire al soggetto destinatario di una decisione automatizzata la possibilità di ottenere informazioni assimilabili, almeno in parte, a una spiegazione di quest'ultima. A questo proposito,

[668] Cfr. ancora, in primo luogo, S. WACHTER, B. MITTELSTADT, L. FLORIDI, *Why a right to explanation of automated decision-making does not exist cit*., p. 90 ss.; v. anche S. RODWAY, *Just How Fair Will Processing Notices Need to Be under the GDPR*, in *Privacy & Data Protection*, 16, 3, 2016.

[669] Lo rilevano, ad esempio, A.D. SELBST, J. POWLES, *Meaningful information and the right to explanation cit.*, che hanno, ad ogni modo, una visione globalmente positiva della formulazione adottata dal GDPR.

[670] In particolare è la tesi principale degli appena citati A.D. SELBST, J. POWLES, *Meaningful information and the right to explanation cit.*

[671] Cfr. B. CASEY, A. FARHANGI, R. VOGL, *Rethinking explainable machines: the GDPR's "right to explanation" debate and the rise of algorithmic audits in enterprise cit.*

[672] ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, 6 febbraio 2018, https://bit.ly/3BaEPTJ (5 agosto 2022).

[673] Cfr. ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Automated individual decision-making cit.*, p. 25.

appaiono particolarmente meritevoli d'analisi le norme di diritto interno francese. Nel 2016, infatti, sono state introdotte in Francia, con la *Loi n 2016-1321 du 7 octobre 2016 pour une République numérique*, alcune importanti garanzie relative alla possibilità per l'individuo di ricevere una spiegazione di eventuali decisioni amministrative assunte per mezzo, anche non esclusivo, di algoritmi.[674]In particolare, l'art. 4 del provvedimento legislativo appena menzionato ha disposto l'introduzione, nel *Code des relations entre le public et l'administration*, di un nuovo articolo L. 311-3-1, il cui primo comma dispone: «une décision individuelle prise sur le fondement d'un traitement algorithmique comporte une mention explicite en informant l'intéressé. Les règles définissant ce traitement ainsi que les principales caractéristiques de sa mise en œuvre sont communiquées par l'administration à l'intéressé s'il en fait la demande»[675]. Un decreto dell'anno successivo ha inserito due ulteriori articoli, che hanno stabilito le caratteristiche delle informazioni da fornire all'interessato della decisione[676]. Ai sensi del nuovo art. L. 311-3-1-1, la *mention* prevista dalla norma precedente deve indicare, almeno, «la finalité poursuivie par le traitement algorithmique» e menzionare il citato diritto di «obtenir la communication des règles définissant ce traitement et des principales caractéristiques de sa mise en œuvre, ainsi que les modalités d'exercice de ce droit»[677]. Secondo il successivo art. L. 311-3-1-2, tale *communication* deve contenere una significativa mole d'informazioni, che probabilmente fanno della norma, allo stato dell'arte, la disposizione più avanzata – e onerosa, anche dal punto di vista della realizzabilità tecnica - in materia di spiegabilità dell'intelligenza artificiale tra quelle attualmente in vigore. Il testo, infatti, impone di fornire «sous une forme intelligible» informazioni riguardanti:«le degré et le mode de contribution du traitement algorithmique à la prise de décision; les données traitées et leurs sources; les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé;les opérations effectuées par le traitement»[678].

---

[674] Loi n 2016-1321 du 7 octobre 2016 *pour une République numérique*, Journal officiel de la République française 235, 8 ottobre 2016. Per dei primi commenti cfr. L. EDWARDS, M. VEALE, *Enslaving the algorithm: from a "right to an explanation" to a "right to better decisions"?*, p. 50 ss.; G. MAUGERI, *Automated decision-making in the EU Member States: the right to explanation and other "suitable safeguards" in the national legislations cit.*, p. 12 ss.

[675] Si riporta la formulazione completa dell'articolo, che prevedeva, al secondo comma, che un decreto ne precisasse le condizioni d'applicazione, come avvenuto l'anno successivo: «sous réserve de l'application du 2° de l'article L. 311-5, une décision individuelle prise sur le fondement d'un traitement algorithmique comporte une mention explicite en informant l'intéressé. Les règles définissant ce traitement ainsi que les principales caractéristiques de sa mise en œuvre sont communiquées par l'administration à l'intéressé s'il en fait la demande. Les conditions d'application du présent article sont fixées par décret en Conseil d'Etat».

[676] Décret n° 2017-330 du 14 mars 2017 *relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique*, Journal officiel de la République française 64, 16 marzo 2017.

[677] Il testo completo dell'articolo è: «La mention explicite prévue à l'article L. 311-3-1 indique la finalité poursuivie par le traitement algorithmique. Elle rappelle le droit, garanti par cet article, d'obtenir la communication des règles définissant ce traitement et des principales caractéristiques de sa mise en œuvre, ainsi que les modalités d'exercice de ce droit à communication et de saisine, le cas échéant, de la commission d'accès aux documents administratifs, définies par le présent livre».

[678] L'art. L. 311-3-1-2, nella sua interezza, recita: «L'administration communique à la personne faisant l'objet d'une décision individuelle prise sur le fondement d'un traitement algorithmique, à la demande de celle-ci, sous une forme

La Francia, inoltre, fa parte della minoranza di paesi europei che hanno identificato degli ambiti di legittimità per i trattamenti consistenti in decisioni totalmente automatizzate ulteriori a quelli previsti dall'art. 22 GDPR, come consentito agli stati membri dell'Unione dal menzionato secondo paragrafo, lett. b) della medesima norma[679]. Come già visto, lo stesso art. 22 GDPR richiede a questi ultimi di prevedere, in tali casi, «misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato»[680]. L'intervento è stato attuato con la *Loi n. 2018-493 du 20 juin 2018 relative à la protection des données personnelles*[681], con cui l'ordinamento francese ha adeguato alle innovazioni introdotte dal GDPR la *Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés*[682]. In seguito alla modifica, l'attuale art. 47 della *Loi informatique* esclude, in primo luogo, che una decisione totalmente automatizzata possa trovare spazio nel sistema giudiziario, al fine di valutare il comportamento o la personalità di un individuo; specifica, in secondo luogo, che una decisione amministrativa totalmente automatizzata deve rispettare gli appena visti criteri introdotti con la *Loi n. 2016-1321 du 7 octobre 2016 pour une République numérique*, e impone, inoltre, al responsabile del trattamento di «s'assure de la maîtrise du traitement algorithmique et de ses évolutions afin de pouvoir expliquer, en détail et sous une forme intelligible, à la personne concernée la manière dont le traitement a été mis en œuvre à son égard»; infine, dispone che qualunque altro trattamento consistente in una decisione automatizzata possa essere messo in atto solamente in virtù delle eccezioni previste dall'art. 22 GDPR, e richiede, come garanzia supplementare, che in tali casi «les règles définissant le traitement ainsi que les principales caractéristiques de sa mise en œuvre soient communiquées, à l'exception des secrets protégés par la loi, par le responsable de traitement à l'intéressé s'il en fait la demande»[683].

---

intelligible et sous réserve de ne pas porter atteinte à des secrets protégés par la loi, les informations suivantes: 1° Le degré et le mode de contribution du traitement algorithmique à la prise de décision; 2° Les données traitées et leurs sources; 3° Les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé; 4° Les opérations effectuées par le traitement».

[679] Per un quadro completo di queste iniziative legislative, il resto delle quali, di portata abbastanza limitata, non prevede misure di particolare interesse ai fini di questo lavoro, cfr. G. MAUGERI, *Automated decision-making in the EU Member States cit.*

[680] L'art. 22 par. 2 lett. b) del GDPR, infatti, prevede che «il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione», previsto dal paragrafo 1, non si applichi qualora il trattamento «sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato».

[681] Loi n. 2018-493 du 20 juin 2018 relative à la protection des données personnelles, Journal officiel de la République française 141, 21 giugno 2018. La modifica dell'art. 47 della *Loi informatique et libertés*, in particolare, è stata attuata con l'art. 7. In letteratura cfr. ancora G. MAUGERI, *Automated decision-making in the EU Member States cit.*, p. 12 ss.

[682] Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, Journal officiel de la République française, 7 gennaio 1978.

[683] L'attuale versione dell'art, 47 della *Loi informatique et libertés*, nella sua interezza, recita: «1. Aucune décision de justice impliquant une appréciation sur le comportement d'une personne ne peut avoir pour fondement un traitement automatisé de données à caractère personnel destiné à évaluer certains aspects de la personnalité de cette personne. 2. Aucune décision produisant des effets juridiques à l'égard d'une personne ou l'affectant de manière significative ne peut être prise sur le seul fondement d'un traitement automatisé de données à caractère personnel, y compris le profilage, à l'exception : 1° Des cas mentionnés aux a et c du 2 de l'article 22 du règlement (UE) 2016/679 du 27 avril 2016, sous les

La legge francese del 2018, dunque, da un lato non pare aggiungere elementi significativi in materia di decisione amministrativa rispetto a quanto già previsto dalla novella di due anni prima, che, come già analizzato, prevede oneri rilevanti in materia di spiegazione del risultato di un sistema. Infatti, la norma esplicita, in modo forse pleonastico, che quanto disposto allora per ogni decisione in cui fosse previsto l'uso di algoritmi, anche come strumento di supporto, vale per le decisioni integralmente demandate alla tecnologia, limitandosi ad aggiungere, in capo al responsabile del trattamento, l'obbligo di «s'assure de la maîtrise du traitement algorithmique et de ses évolutions», peraltro di difficile realizzabilità, in mancanza di ingenti iniziative volte alla formazione specifica dei funzionari pubblici. Dall'altro lato, l'intervento riformatore prevede innovazioni di rilievo riguardo ogni altra decisione automatizzata, presa da attori pubblici e privati, chiarendo che, in ogni caso, il destinatario di tali decisioni ha il diritto di conoscere «les règles définissant le traitement ainsi que les principales caractéristiques de sa mise en œuvre». Per la verità, la disposizione appare piuttosto ermetica, e non in grado di risolvere definitivamente i dubbi sull'estensione delle garanzie individuali in materia di spiegazione dell'intelligenza artificiale, esponendosi al rischio di un'interpretazione restrittiva che finisca per considerarla una norma meramente riepilogativa dell'obbligo di fornire all'interessato «informazioni significative sulla logica utilizzata», come visto già previsto dal GDPR. Ciò nonostante, con la sua introduzione l'ordinamento francese fa di certo un passo avanti verso il riconoscimento di un pieno diritto individuale alla spiegazione, con una norma di diritto interno utile a non far dipendere il riconoscimento di quest'ultimo dal solo diritto europeo, la cui formulazione in materia, come già visto, appare particolarmente incerta e ambigua.

Al di fuori del contesto europeo, il principale esempio di una possibile positivizzazione del diritto a una spiegazione è rappresentato, ancora una volta, dalla *Directive on automated decision-making* canadese, il cui par. 6.2.3 impone alle pubbliche amministrazioni che si avvalgano di algoritmi di fornire «a meaningful explanation to affected individuals of how and why the decision was made as prescribed in Appendix C»[684]. Tale appendice alla direttiva distingue le caratteristiche di tale

réserves mentionnées au 3 du même article 22 et à condition que les règles définissant le traitement ainsi que les principales caractéristiques de sa mise en œuvre soient communiquées, à l'exception des secrets protégés par la loi, par le responsable de traitement à l'intéressé s'il en fait la demande; 2° Des décisions administratives individuelles prises dans le respect de l'article L. 311-3-1 et du chapitre Ier du titre Ier du livre IV du code des relations entre le public et l'administration, à condition que le traitement ne porte pas sur des données mentionnées au I de l'article 6 de la présente loi. Ces décisions comportent, à peine de nullité, la mention explicite prévue à l'article L. 311-3-1 du code des relations entre le public et l'administration. Pour ces décisions, le responsable de traitement s'assure de la maîtrise du traitement algorithmique et de ses évolutions afin de pouvoir expliquer, en détail et sous une forme intelligible, à la personne concernée la manière dont le traitement a été mis en œuvre à son égard. 3. Par dérogation au 2° du présent article, aucune décision par laquelle l'administration se prononce sur un recours administratif mentionné au titre Ier du livre IV du code des relations entre le public et l'administration ne peut être prise sur le seul fondement d'un traitement automatisé de données à caractère personnel».

[684] La norma è compresa tra gli obblighi di trasparenza di varia natura previsti dal paragrafo 6.2, per l'appunto rubricato *Transparency*, che impone, tra le altre cose, di informare l'interessato dell'esistenza del sistema automatizzato (par. 6.2.1-2) e di assicurare che l'Amministrazione conservi comunque la possibilità di esercitare un controllo sul sistema,

spiegazione in base al rischio connesso alla tecnologia utilizzata. In particolare, è identificata una categoria di decisioni considerate a basso rischio, in ragione del loro potenziale lesivo nullo o limitato – l'Appendix B, che definisce i livelli di rischio, parla di «little or no impact» (Level I) - per alcuni interessi primari: «the rights of individuals or communities; the health or weel-being of individuals or communities; the economic interest of individuals, entities or communities; the ongoing sustainability of an ecosystem»[685]. In questi casi, la spiegazione può consistere in semplici informazioni sulle più comuni modalità di funzionamento del sistema, senza alcun collegamento con la singola decisione in esame. L'Appendix C, infatti, nel descrivere le modalità di adempimento all'*explanation requirement* previsto dalla direttiva, si esprime in questi termini: «in addition to any applicable legal requirement, ensuring that a meaningful explanation is provided for common decision results. This can include providing the explanation via a Frequently Asked Questions section on a website.»[686]. Invece, per ogni altra decisione – e dunque in tutti i casi in cui il possibile impatto sugli interessi già visti non appaia lieveo inesistente – è previsto, in capo all'autorità che si avvalga di algoritmi, il dovere di garantire al soggetto interessatouna spiegazione specifica in tutti i casi in cui tale decisione consista nel rifiuto di un beneficio o servizio, o sia comunque assimilabile all'esercizio di un potere (il testo parla di «regulatory action»)[687]. Tuttavia, la formulazione della norma è piuttosto ermetica, e non offre spunti concreti riguardo alle informazioni che dovranno essere fornite con tale spiegazione, esponendosi al rischio di interpretazioni estremamente riduzioniste, potenzialmente in grado di annullarne il portato garantistico. Un pericolo, come già visto, condiviso con le norme eurounitarie sulla stessa materia, con le quali, del resto, la direttiva canadese condivide anche alcune scelte terminologiche (in primo luogo l'impiego del termine *meaningful* come unico connotato della spiegazione, tradotto con «significative» nella versione italiana del GDPR). Il citato Annex C, infatti, stabilisce che, per i casi di rischio non minimo, l'*explanation requirement* debba essere adempiuto «ensuring that a meaningful explanation is provided with any decision that resulted in the denial of a benefit, a service, or other regulatory action».

---

anche esaminandone le singole componenti, qualora siano utilizzati strumenti coperti da segreto industriale (par. 6.2.4-5).

[685] Come già riportato, l'Appendix B distingue quattro classi di rischio, in base al rischio per i beni e valori menzionati nel testo: *little or no impact* (Level I): «Level I decisions will often lead to impacts that are reversible and brief»; *moderate impacts* (Level II): «will often lead to impacts that are likely reversible and short-term»; *high impacts* (Level III): «Level III decisions will often lead to impacts that can be difficult to reverse, and are ongoing»; *very high impacts* (Level IV): «will often lead to impacts that are irreversible, and are perpetual».

[686] Questo, infatti, è quanto dispone l'Appendix C per le decisioni classificate al Level I della suddivisione in fasce di rischio operata dall'Appendix B, di cui alla nota precedente.

[687] Per le decisioni classificabili ai livelli di rischio II, III o IV definiti dall'Appendix B, infatti, l'Appendix C, per quanto riguarda l'*explanation requirement*, impone: «in addition to any applicable legal requirement, ensuring that a meaningful explanation is provided with any decision that resulted in the denial of a benefit, a service, or other regulatory action».

3.5 La spiegabilità dei sistemi come diritto fondamentale: le lacune del quadro giuridico esistente

Come già anticipato, l'inclusione della garanzia di una spiegazione dei risultati di un sistema tra i diritti fondamentali ha incontrato con facilità riconoscimento nella letteratura giuridica[688]. Ciò in ragione del particolare ruolo rivestito dall'idea di spiegazione per un'ampia gamma di diritti fondamentali[689]. La possibilità di comprendere le ragioni essenziali di una determinata decisione, infatti, rappresenta, come già detto, il presupposto logico di ogni garanzia riconosciuta dal costituzionalismo nei rapporti tra individuo e potere[690]. Una determinata motivazione – spesso imposta con norme di rango costituzionale - deve sussistere, ad esempio, per la limitazione, da parte dei poteri pubblici, di libertà individuali fondamentali, come il domicilio e la corrispondenza; ogni provvedimento dell'autorità giudiziaria o amministrativa in grado di incidere sulla sfera giuridica dei privati deve chiarire le sue ragioni[691]; garanzie procedurali estremamente rigorose – si pensi, in primo luogo, al processo penale – sono costruite al fine di individuare le cause di determinati fatti da cui l'ordinamento fa discendere conseguenze per i diritti individuali, permettendo in ogni momento all'individuo interessato di intervenire presentando spiegazioni alternative[692]. In più, l'identificazione dei motivi delle proprie e altrui azioni e la ricostruzione di legami eziologici tra gli eventi fanno parte del modo con cui l'essere umano organizza e comunica le proprie conoscenze sul mondo[693]. Ciò è vero, pur con importanti differenze, pressoché in ogni cultura, e a partire dalle

---

[688] Cfr. ad esempio M. WINIKOFF, J. SARDELIC, *Artificial Intelligence and the Right to Explanation as a Human Right*, in *IEEE Internet Computing*, 25, 2, 2021, p. 116-120 e ancora C. CASONATO, *Intelligenza artificiale e dirtto costituzionale: prime considerazioni cit.*; A. SIMONCINI, *L'algoritmo incostituzionale cit.*; L.M. AZZENA, *L'algoritmo nella formazione della decisione amministrativa cit.*

[689] Si vedano, ad esempio, M. RISSE, *Human Rights and Artificial Intelligence: An Urgently Needed Agenda*, in *HKS Faculty Research Working Paper Series RWP18-015*, 2018; ACCESS NOW, *Human rights in the age of artificial intelligence-Report*, 2018, https://bit.ly/2GiFWGT (6 agosto 2022). Cfr. anche l'impostazione di A.D. SELBST, J. POWLES, *Meaningful information and the right to explanation cit.* secondo il quale il contenuto minimo delle informazioni da fornire all'interessato per soddisfare i requisiti imposti dal GDPR in materia di spiegazione deve consentire, appunto, di riconoscere, ed eventualmente portare di fronte a un giudice, di aver subito violazioni dei diritti fondamentali, in primo luogo in punto di non discriminazione.

[690] È significativo notare, sul punto, come il ruolo fondamentale della motivazione nel discorso giuridico, quale elemento di razionalità e comprensibilità in contesti caratterizzati da incertezza e forti scontri valoriali, sia stato al centro delle riflessioni dei maggiori teorici del diritto della seconda parte del XX secolo, cfr. in particolare R. ALEXI, *Teoria dell'argomentazione giuridica. La teoria del discorso razionale come teoria della motivazione giuridica*, Milano, 1988; hanno rivestito un ruolo decisivo, in tale percorso, anche i filosofi della c.d. svolta argomentativa, cfr. ad es. C. PERELMAN, L. OLBRECHTS-TYTECA, *Traité de l'argumentation, la nouvelle rhétorique*, Bruxelles, 1958; S. TOULMIN, *The uses of argument*, Cambridge, 1958.

[691] Sul ruolo della motivazione nei provvedimenti giudiziali e amministrativi, tra i numerosissimi studi sul tema si rimanda, a mero titolo esemplificativo, ai noti R. JUSO, *Motivi e motivazione nel provvedimento amministrativo*, Milano, 1960; M. TARUFFO, *La motivazione della sentenza civile*, Padova. 1975.

[692] L'inesauribile dibattito dottrinale e giurisprudenziale sulla causalità nel diritto penale ben illustra il ruolo centrale rivestito, in tale ambito, dalla spiegazione. *Ex multis*, si rimanda agli studi in materia di F. STELLA, *Leggi scientifiche e spiegazione causale nel diritto penale*, Milano, 1990.

[693] Ciò è vero, in primo luogo, nelle culture occidentali, il cui modello di razionalità è fortemente connesso allo sviluppo del metodo scientifico. L'antropologia, comunque, ha messo in luce come la ricerca di una spiegazione ad azioni e

origini della specie umana, tanto che vi sono tracce dello sviluppo di un pensiero religioso, ovviamente connesso alla necessità di interpretare elementi della realtà non spiegabili sul piano razionale, fin da epoche antichissime[694]. L'uomo, dunque, non ha mai pacificamente accettato di non poter comprendere la realtà che lo circonda e questo, del resto, è sempre stato il principale motore verso l'acquisizione di nuove conoscenze. Un paradigma consolidato che si trova inaspettatamente messo in discussione da innovazioni tecnologiche che sembrano sottrarsi alla possibilità di essere interpretate con gli strumenti della razionalità. È chiaro, allora, che la messa in dubbio di una prerogativa talmente basilare da riguardare il modo in cui l'umano ha sempre concepito la sua relazione con il mondo esterno abbia portato a riconoscere con facilità che la garanzia di spiegazione dei risultati di tali tecnologie appartenesse al campo dei diritti fondamentali. Elevare la spiegazione al rango di diritto fondamentale è decisivo anche per giudicare dalla giusta prospettiva le prime iniziative per una sua positivizzazione, analizzate nelle pagine precedenti. Le critiche mosse all'ambiguità delle formule legislative scelte appaiono almeno in parte condivisibili, nonostante non possa trascurarsi che le prime interpretazioni della norma, e in particolare, sullo scenario europeo, quella dello EDPB[695], sembrino andare verso una sua valorizzazione, e non verso un restringimento del suo ambito applicativo. Allo stesso tempo, non può trascurarsi nemmeno che, allo stato dell'arte, le parole «diritto alla spiegazione» compaiano solamente al Cons. 71 del GDPR. Per quanto tale riconoscimento sia senza dubbio estremamente significativo, prima di tutto a livello simbolico (la spiegazione è, appunto, definita un diritto, e non trattata alla stregua di un semplice requisito tecnico) si deve evidenziare che tale disposizione non ha, nel diritto europeo, pieno valore vincolante. Le lacune dell'assetto normativo predisposto dall'Unione Europea, inoltre, non possono bilanciarsi con la circostanza che il GDPR preveda, allo stesso tempo, strumenti di *enforcement*di inedita intensità, come pur qualcuno ha teorizzato[696]. Per quanto tali strumenti siano senza dubbio significativi, infatti, l'efficacia delle strategie volte ad assicurare l'effettività del diritto non può portare a cambiare il giudizio sul suo contenuto: lo dimostra che esse potrebbero essere impiegate per finalità del tutto opposte, se mutassero le norme di cui presidiano l'applicazione. Se è certo che un diritto non possa dirsi pienamente positivizzato senza la predisposizione di idonei mezzi di

fenomeni e l'organizzazione del mondo in categorie concettuali e linguistiche appartengano ad ogni cultura e si siano, presumibilmente, sviluppati fin dalle epoche più remote, al netto del noto dibattito sulla possibilità, o meno, di identificare degli elementi comuni tra i distinti sistemi di comprensione e comunicazione delle conoscenze, cfr. ad es. C.W. NUCKOLLS, *The Anthropology of Explanation*, in *Anthropological Quarterly*, 66, 1, 1993, p. 1 ss.; S. J. TAMBIAH, *Magic, science, religion, and the scope of rationality*, Cambridge, 1990; R. HORTON, *African Traditional Thought and Western Science*, in *Africa*, 37, 1, 1967, p. 50-71.

[694] Cfr. ad es. E. CULOTTA, *On the Origin of Religion*, in *Science*, 326, 5954, 2009, p. 784-787; R. STARK, *Discovering God: the origins of the great religions and the evolution of belief*, New York, 2007, p. 10 ss.

[695] Si fa riferimento alle citate ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Automated individual decision-making cit.*, p. 25.

[696] Cfr. ad es. B. CASEY, A. FARHANGI, R. VOGL, *Rethinking explainable machines: the GDPR's "right to explanation" debate and the rise of algorithmic audits in enterprise cit.*

enforcement a supporto, allora, è altrettanto certo che non sono quest'ultimi a determinarne il riconoscimento in un ordinamento[697].

Allo stesso tempo, preme evidenziare come la principale mancanza delle norme con cui il diritto europeo ha introdotto alcuni requisiti in materia di spiegabilità dell'intelligenza artificialenon stia nell'ambiguità del testo del GDPR, ma nell'ambito di applicazione che esso individua. Infatti, l'art. 22 GDPR, come già più volte ricordato, circoscrive il suo raggio d'azione ai trattamenti che consistano in decisioni totalmente automatizzate. Come osservato in letteratura, non pare razionale la scelta di limitarne gli effetti alle decisioni interamente svolte dagli algoritmi, poiché ciò porta all'esclusione di sistemi di supporto alla decisione dell'essere umano di grande rilievo pratico e per i quali è agevole prevedere un rapido sviluppo nei prossimi anni e un altrettanto rapida diffusione sul mercato, anche in contesti che pongono questioni etiche di intriseca delicatezza, come l'ambito medico[698].I limiti dell'impostazione del GDPR, però, non si esauriscono di certo qui. In primo luogo, infatti, la norma, come l'intero Regolamento, prende in considerazione i soli trattamenti di dati personali. In tal modo, resta esclusa dalle garanzie da essa previste ogni applicazione tecnologica che utilizzi, per elaborare i propri risultati, dati non rientranti nella definizione legislativa di dato personale[699]. La circostanza non implica conseguenze di poco conto, posto che

[697] Sulla distinzione tra diritti e loro garanzie cfr. ancora L. FERRAJOLI, *Diritti fondamentali cit.*, p. 9 ss.

[698] A rilevarlo, in particolare, sono L. EDWARDS, M. VEALE, *Enslaving the algorithm cit.*,p. 3. Per le applicazioni dell'intelligenza artificiale a supporto della decisione medica, cfr., tra gli altri, Deloitte, *The future of AI in healthcare. How AI will impact patients, clinicians and the pharmaceutical industry*, https://bit.ly/3Ql9qmM, 2019; I. KICKBUSCH, D. PISELLI, A. AGRAWAL, R. BALICER, O. BANNER, M. ADELHARDT ET AL., *The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world*, in *Lancet*, 398, 10312, 2021, p. 1727-1776. Si rinvia, inoltre, all'analisi del tema svolta *infra*, p. 277 ss.

[699] Per alcune ricerche, relative allo sviluppo di modelli predittivi, in cui l'utilizzo di dati anonimizzati sia posto particolarmente in enfasi, cfr. ad esempio H.H. ARCOLEZI, J.F. COUCHOT, S. CERNA, C. GUYEUX, G. ROYER, B.A. BOUNA ET AL., *Forecasting the number of firefighter interventions per region with local-differential-privacy-based data*, in *Computers & Security*, 96, 2020, p. 101888 ss.; Y. FENG, Q. DUAN, X. CHEN, S.S. YAKKALI, J. WANG, *Space cooling energy usage prediction based on utility data for residential buildings using machine learning methods*, in *Applied Energy*, 291, 2021, p. 116814 ss.; C.C. CHANG, B. THOMPSON, H. WANG, C.G. NESPEREIRA, E. ELHARIRI, N. EL-BENDARY, A.F. VILAS, R.P.D. REDONDO, *Machine Learning Based Classification Approach for Predicting Students Performance in Blended Learning*, in T. GABER, A. E. HASSANIEN, N. EL-BENDARY, N. DEY (A CURA DI), *The 1st International Conference on Advanced Intelligent System and Informatics (AISI2015)*, Cham, 2016, p. 47-56. Deve evidenziarsi, in ogni caso, che molti dei sistemi predittivi basati sull'elaborazione di grandi molti di dati con tecnologie di intelligenza artificiale sono allenati con *database* di dati non personali o anonimizzati, o che comunque sarebbe possibile anonimizzare (la normativa in materia di dati personali concretamente applicata, ovviamente, incide sull'attenzione a questo profilo). Il trattamento di dati personali avviene al momento di utilizzare il sistema sviluppato in tal modo per decisioni e valutazioni attinenti a una persona fisica specificamente indviduata. Qualora, però, il database sia utilizzato per scopi differenti, come la prognosi di tendenze generali economiche o demografiche è ipotizzabile che l'intera attività risulti al di fuori dell'ambito di applicazione del GDPR. Sulla difficoltà a stabilire gli effettivi confini della stessa categoria dei dati anonimizzati, e sul continuo rincorrersi tra strategie di crittografia e anonimizzazione sempre più elaborate e sistemi di reidentificazione altrettanto innovativi, si rimanda ai già citati L. SWEENEY, M. VON LOEWENFELDT, M. PERRY, *Saying it's anonymous doesn't make it so: re-identification of "anonymized" law school data*, in *Technology Science*, November 12, 2018, https://techscience.org/a/2018111301/ (7 marzo 2021); L. SWEENEY, *Only you, your doctor and many others may know*, in *Technology Science*, September 28, 2015, https://techscience.org/a/2015092903/ (7 marzo 2021); A. J. COHEN, *New guarantees for cryptographic circuits and data anonymization*, Cambridge, 2019, p. 235 ss.; A. ANTONIOU, G. DOSSENA, J. MACMILLAN, S. HAMBLIN, D.

l'analisi su larga scala di dati anonimizzati può comunque essere utilizzata per attività in senso lato decisionali, in grado di dispiegare importanti effetti indiretti sulla vita del singolo o di intere classi di individui. Si pensi, per limitarsi a una sola ipotesi, alla modifica delle strategie di investimento pubblico su scala pluriennale in determinate misure di welfare in funzione di valutazioni algoritmiche dell'efficacia su larga scala di quese ultime in passato, e all'eventualità che le tecnologie coinvolte si comportino come *black-box*, senza che nessuno sollevi il problema[700]. In secondo luogo, i contesti in cui può, potenzialmente, venire in rilievo la necessità di ottenere una spiegazione dell'*output* di un sistema intelligente non sono, di certo, limitati all'ambito dell'attività decisionale, totalmente o parzialmente automatizzata, che riguardi direttamente uno o più individui determinati. L'intelligenza artificiale, infatti, è utilizzata in un gran numero di settori non riconducibili a tale contesto se non con importanti sforzi interpretativi e in cui l'eventuale utilizzo di tecnologie non interpretabili pone di certo interrogativi urgenti. Si pensi, per limitarsi solo a due esempi, alla grande robotica industriale impiegata in contesti produttivi ibridi e ai rischi per l'incolumità dei lavoratori che possono conseguire da errori o anomalie di funzionamento[701], o a sistemi di intelligenza artificiale impiegati in attività intrinsecamente pericolose come il trasporto aereo.

3.6 Le prospettive dischiuse dalla Proposta di Regolamento europeo sull'intelligenza artificiale

In breve, le norme in materia di spiegabilità dell'intelligenza artificiale di recente introdotte nell'Unione Europea e in altri ordinamenti non sembrano in grado di assicurare la tutela del diritto alla spiegazione in ogni campo in cui esso potrebbe rappresentare un'efficace difesa dell'individuo, come dovrebbe accadere per una posizione giuridica che, secondo la prospettiva adottata in questo lavoro, si eleva al rango di diritto fondamentale. In ogni caso, la Proposta di Regolamento dell'Unione Europea dell'aprile 2021 in materia di intelligenza artificiale pare farsi carico, almeno

---

CLIFTON, P. PETRONE, *Assessing the risk of re-identification arising from an attack on anonymized data*, 2022, arXiv:2203.16921, cfr. inoltre *supra*, p. 96-98.

[700] Il riferimento obbligato è alle riflessioni di G. CALABRESI, P. BOBBITT, *Tragic Choices*, New York-Londra, 1978, che distinguono, nell'allocazione di risorse scarse di estrema importanza, due classi di decisioni: le *second-order choices*, ovvero le decisioni tragiche comunemente identificate, volte a definire il destinatario di un determinato beneficio, ad esempio un trattamento di sostegno vitale, e le *first-order choices*, ossia le scelte che, a monte, determinano tale scarsità, definendo, ad esempio con la programmazione economica, la quantità di risorse che sarà concretamente disponibile. I due autori rilevano come la tragicità di queste ultime, le gerarchie tra valori che sottendono e, talvolta, la loro stessa esistenza e influenza su ogni atomistica *second-order choice* passi spesso inosservata, sia tra il pubblico che tra gli addetti ai lavori. L'analisi prende a riferimento un mondo in larga parte predigitale, ma risulta di grande attualità: il legame tra scelte di primo e secondo ordine, infatti, riguarda qualunque politica di welfare, e l'eventuale coinvolgimento di strumenti di intelligenza artificiale rischia, come detto nel testo, di aumentare ulteriormente tale difficoltà di percezione della moralità delle *first-order choice*. I giudizi di valore che implicano, infatti, sarebbere nascosti dall'apparenza di oggettività che frequentemente accompagna la tecnologia e dall'eventuale opacità del suo funzionamento.

[701] Per i potenziali rischi per l'operatore umano cfr. A.F.T. WINFIELD ET AL., *Robot Accident Investigation: A Case Study in Responsible Robotics*, in A. CAVALCANTI, B. DONGOL, R. HIERONS, J. TIMMIS, J. WOODCOCK (A CURA DI), *Software Engineering for Robotics*, Cham, 2021, p. 165 ss.

in parte, del problema, con una disciplina che identifica un'ambito di applicazione molto più ampio delle norme analizzate finora e si esprime, al contempo, in modo molto più chiaro e risoluto. L'art. 13 della Proposta, infatti, richiede che i sistemi di IA ad alto rischio (il cui ambito, come visto nella prima parte del lavoro, è comunque molto ampio)[702] siano «progettati e sviluppati in modo tale da garantire che il loro funzionamento sia sufficientemente trasparente da consentire agli utenti di interpretare l'output del sistema e utilizzarlo adeguatamente. Sono garantiti un tipo e un livello di trasparenza adeguati, che consentano di conseguire il rispetto dei pertinenti obblighi dell'utente e del fornitore di cui al capo 3 del presente titolo [obblighi di condotta, verifica e *compliance* che assicurano l'effettiva applicazione delle garanzie previste per i sistemi ad alto rischio, ndr]».[703]Com'è evidente, la norma è di natura tecnica, tanto che i requisiti di trasparenza da essa previsti sono concepiti come un ausilio all'attività dell'utente del sistema e uno strumento per il rispetto degli obblighi che il Regolamento pone al fornitore della tecnologia utilizzata, senza menzionare il destinatario degli effetti degli *output* di quest'ultima, che potrebbe essere persona diversa dall'utente. L'ampio campo d'applicazione e la chiarezza del precetto, portano, però, a ipotizzare che essa, nel caso di effettiva approvazione ed entrata in vigore, potrebbe rappresentare un efficace presidio, ad ampio spettro, degli interessi connessi alla spiegabilità dei sistemi, visti nelle pagine precedenti. Dunque, nonostante la formulazione meno altisonante rispetto al GDPR – nella Proposta di Regolamento non vi è traccia di espressioni come "diritto alla spiegazione", come visto presente, invece, nel Cons. 71 del Regolamento in materia di dati personali –l'*Artificial*

[702] È l'art. 6 della Proposta di Regolamento, come già evidenziato, a individuare le applicazioni dell'intelligenza artificiale ad alto rischio. La norma indica un'ampia serie di tecnologie, con un sistema di rinvii agli allegati del testo legislativo che risulta di non facile lettura. Sono applicazioni ad alto rischio tutte quelle elencate all'allegato III (che comprende, ad esempio, i sistemi utilizzati nell'attività creditizia, la gestione dei flussi migratori, o per l'identificazione biometrica nei casi in cui il precedente art. 5 la permette) e tutti i sistemi di intelligenza artificiale che siano prodotti o componenti di sicurezza di prodotti disciplinati dalla normativa di armonizzazione indicata all'allegato II, per i quali sia prevista una valutazione di conformità da parte di terzi. Cfr. *supra*, p. 69 ss.

[703] Così recita il primo paragrafo dell'articolo, mentre quelli successivi sono dedicati alle informazioni che devono essere fornite col sistema di intelligenza artificiale, riguardanti le modalità di realizzazione dei molteplici requisiti che la Proposta prevede per le tecnologie ad alto rischio: «2. I sistemi di IA ad alto rischio sono accompagnati da istruzioni per l'uso in un formato digitale o non digitale appropriato, che comprendono informazioni concise, complete, corrette e chiare che siano pertinenti, accessibili e comprensibili per gli utenti. 3. Le informazioni di cui al paragrafo 2 specificano: a) l'identità e i dati di contatto del fornitore e, ove applicabile, del suo rappresentante autorizzato; b) le caratteristiche, le capacità e i limiti delle prestazioni del sistema di IA ad alto rischio, tra cui: i) la finalità prevista; ii) il livello di accuratezza, robustezza e cibersicurezza di cui all'articolo 15 rispetto al quale il sistema di IA ad alto rischio è stato sottoposto a prova e convalidato e che ci si può attendere, e qualsiasi circostanza nota e prevedibile che possa avere un impatto sul livello atteso di accuratezza, robustezza e cibersicurezza; iii) qualsiasi circostanza nota o prevedibile connessa all'uso del sistema di IA ad alto rischio in conformità alla sua finalità prevista o in condizioni di uso improprio ragionevolmente prevedibile, che possa comportare rischi per la salute e la sicurezza o per i diritti fondamentali; iv) le sue prestazioni per quanto riguarda le persone o i gruppi di persone sui quali il sistema è destinato a essere utilizzato; v) ove opportuno, le specifiche per i dati di input o qualsiasi altra informazione pertinente in termini di set di dati di addestramento, convalida e prova, tenendo conto della finalità prevista del sistema di IA; c) le eventuali modifiche apportate al sistema di IA ad alto rischio e alle sue prestazioni, che sono state predeterminate dal fornitore al momento della valutazione iniziale della conformità; d) le misure di sorveglianza umana di cui all'articolo 14, comprese le misure tecniche poste in essere per facilitare l'interpretazione degli output dei sistemi di IA da parte degli utenti; e) la durata prevista del sistema di IA ad alto rischio e tutte le misure di manutenzione e cura necessarie per garantire il corretto funzionamento di tale sistema, anche per quanto riguarda gli aggiornamenti software».

Intelligence Act europeo sembra poter rappresentare, qualora approvato nella versione attuale, un importante indice del riconoscimento, da parte dell'ordinamento europeo, della garanzia di una spiegazione come un vero e proprio diritto fondamentale.

3.7 Un diritto alla spiegazione è veramente possibile? La sostenibilità tecnologica della trasparenza algoritmica e il suo ruolo nel bilanciamento con altri diritti e interessi

Esaurita la disamina delle prospettive di positivizzazione del diritto a una spiegazione, rimane aperto il tema della sostenibilità tecnologica di quest'ultimo. Lo stato dell'arte in materia di sistemi c.d. *black-box* ed *explainable artificial intelligence* è stato analizzato ai paragrafi precedenti, dando conto di come, in estrema sintesi, una piena soluzione del problema, dal punto di vista tecnologico, non sia ancora ipotizzabile[704]. La prima variabile da analizzare riguarda, come più volte ripetuto, il concetto di spiegazione che si intende prendere a riferimento. Da questo punto di vista, la grande varietà delle applicazioni tecnologiche che potrebbero venire in esame e dei loro potenziali scenari applicativi impone di non aderire a una singola concezione di spiegazione: ciascuna delle strategie di *explainable artificial intelligence* già viste può, in astratto, rappresentare un'idonea protezione del diritto alla spiegazione. Le norme analizzate sembrano procedere nella stessa direzione, utilizzando formulazioni generiche, adattabili alle diverse possibili strategie per ottenere una motivazione dei risultati di un sistema. Tale ultima circostanza sembra valere anche per l'art. 13 della Proposta di Regolamento dell'Unione Europea attualmente in discussione, che, come già riportato, richiede che le tecnologie di intelligenza artificiale siano«sufficientemente trasparent[i] da consentire agli utenti di interpretare l'output». La sintetica espressione "interpretare l'*output*" (requisito, come si vedrà al capitolo seguente, ribadito anche dal successivo art. 14)[705] infatti, non pare esprimere una preferenza verso determinate tecniche di *explainable artificial intelligence*[706].
Saranno, dunque, le circostanze del caso concreto a stabilire quale sia la tipologia di spiegazione più idonea a garantire il rispetto del diritto alla spiegazione. Non possono trascurarsi, però, i già visti

[704] Si rimanda, da vari punti di vista, ai già citatiS. KRISHNA, T. HAN, A. GU, J. POMBRA, S. JABBARI, S. WU, *The Disagreement Problem in Explainable cit.*;W. HRYNIEWSKA, P. BOMBIŃSKI, P. SZATKOWSKI ET AL., *Do not repeat these mistakes cit.*; Z. BUÇINCA, P. LIN, K.Z. GAJOS, E.L. GLASSMAN, *Proxy tasks and subjective measures can be misleading cit.*;H. DE BRUIJN M. WARNIER, M. JANSSEN, *The perils and pitfalls of explainable AI cit.*; A. DRAGAN, M. HARDT, *Model Reconstruction from Model Explanationscit.*; A. BURT, *The AI transparency paradox cit.*R.V. YAMPOLSKIY, *Unexplainability and incomprensibility cit.*; A. ADADI, M. BERRADA,*Peeking Inside the Black-Box cit.*; S. SARKAR ET AL., *Accuracy and Interpretability Trade-offs in Machine Learning Applied to Safer Gambling cit.*

[705] Come si vedrà *infra*, p. 224 ss., l'art. 14 della Proposta di Regolamento elenca una minuziosa serie di requisiti che i sistemi intelligenti dovranno rispettare in materia sorveglianza umana, uno dei quali (art. 14 par. 4 lett. c) impone di garantire che lo *human in the loop* sia messo in grado «di interpretare correttamente l'output del sistema di IA ad altorischio, tenendo conto in particolare delle caratteristiche del sistema e degli strumenti e dei metodi di interpretazione disponibili».

[706] L'espressione, infatti, può indicare sia strategie volte a comprendere gli stati interni della tecnologia *black-box* (la formulazione appropriata per restringere il campo a queste ultime sarebbe stata, infatti, interpretare *il sistema*, e non *l'output*) che approcci basati sulla costruzione di spiegazioni *ex post* dei suoi risultati.

limiti che, allo stato dell'arte, il settore dell'*explainable artificial intelligence* presenta. Alcune spiegazioni potrebbero avere un valore euristico limitato, apparire come semplici ipotesi elaborate *ex post* senza alcun collegamento con la tecnologia di riferimento, o fornire informazioni che a taluni appaiono insufficienti (può essere il caso, ad esempio, delle tecniche di *explainable artificial intelligence* basate sul ragionamento controfattuale)[707]. Per definire che genere di spiegazione sia preferibile, o possa dirsi comunque sufficiente, sarà necessario guardare ai valori di volta in volta gioco. Il diritto fondamentale alla spiegazione dovrà, allora, essere bilanciato con gli altri diritti e interessi coinvolti in una determinata situazione, al pari di ogni altro diritto fondamentale. Potranno esservi, così, situazioni in cui anche una spiegazione minima potrà ritenersi sufficiente, per la rilevanza, massima o estremamente ridotta, dei valori coinvolti. Come già analizzato nei precedenti paragrafi, può essere il caso, rispettivamente, dell'utilizzo di determinate tecnologie c.d. *black-box* in ambito sanitario, in grado di rappresentare una protezione del diritto alla vita e all'integrità fisica non raggiungibile, allo stato dell'arte, con sistemi interpretabili[708], o di applicazioni dell'intelligenza artificiale operanti in contesti in cui la protezione del diritto alla spiegazione non sia concepita come necessaria nemmeno dai suoi astratti titolari[709]. D'altro canto, potranno esservi contesti in cui il bilanciamento imponga il pieno rispetto del diritto alla spiegazione, fino alla radicale rinuncia a tecnologie non pienamente interpretabili, a prescindere dalla loro eventuale maggiore efficienza. Un valido esempio può essere rappresentato dalla decisione giudiziaria, che difficilmente pare poter essere delegata, o anche semplicemente supportata, da algoritmi *black-box*, senza snaturare l'essenza delle garanzie processuali alla base della nostra tradizione giuridica e di quelle ad essa più affini[710]. Le conseguenze in materia di diritti fondamentali dell'utilizzo di sistemi di intelligenza artificiale nei settori di sanità e giustizia saranno, in ogni caso, analizzate più approfondimentamente nel prossimo capitolo. Ciò che preme qui evidenziare, invece, è che l'applicazione pratica di un eventuale diritto a una spiegazione dei risultati di un sistema sarà determinata dal bilanciamento con gli altri diritti e interessi primari coinvolti, al pari di ogni altro diritto fondamentale. Tale nuova posizione giuridica, però, sarebbe idonea, in tutti i casi, a stimolare

[707] Si rimanda, in primo luogo, al citato T. MILLER, *Explanation in artificial intelligence cit.*. Deve ribadirsi che, in ogni caso, ogni genere di spiegazione, a prescindere dal coinvolgimento di tecnologie avanzate, consiste in un'approssimazione, cfr. ad esempio J. C. PITT (A CURA DI), *Theories of explanation cit.*; W. G. LYCAN, *Explanation and epistemology cit.*

[708] Un esempio, come già indicato, può essere rappresentato da determinate tecnologie di supporto alla diagnosi basate sull'*image recognition*, in cui l'utilizzo di reti neurali profonde ha portato a risultati considerati equivalenti o migliori a quelli di sanitari esperti, cfr. A. HEKLER ET AL., *Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images*, in *European Journal of Cancer*, 118, 2019, p. 91 ss.; A. ESTEVA, B. KUPREL, R. A. NOVOA, J. KO, S. M. SWETTER, H. M. BLAU, *Dermatologist-level classification of skin cancer with deep neural networks cit.*

[709] Si rimanda, in primo luogo, al già citato studio di A. BUNT, M.LOUNT, C. LAUZON, *Are explanations always important? cit.*

[710] Cfr. ad esempio C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi*, in *DPCE Online*, 44, 3, 2020, p. 379 ss. Il tema sarà analizzato più in profondità *infra*, p. 256 ss.

l'implementazione delle più avanzate tecniche di *explainable artificial intelligence* esistenti, stimolerebbe ricerca e industria verso ulteriori sviluppi in materia e rappresenterebbe una fondamentale garanzia dell'utilizzo di sistemi *black-box* solamente in presenza di ragioni effettive, in primo luogo in termini di efficacia, che sconsiglino l'uso di sistemi interpretabili. Il diritto in esame dovrebbe, in sintesi, fare sì che sia, in ogni caso, adottata ogni strategia al fine di ottenere il maggior livello di spiegabilità possibile[711].

4. Il diritto al controllo umano sul sistema o *human in the loop*

4.1. Il rapporto tra essere umano e automazione e il mutamento di paradigma conseguente all'avvento dell'intelligenza artificiale

L'automazione di un numero sempre crescente di attività – non necessariamente per mezzo dell'intelligenza artificiale – solleva il tema del ruolo residuo dell'essere umano in tali contesti, dei quali, tipicamente, era il protagonista prima dell'avvento di un determinato sviluppo tecnologico. In termini generali, la funzione dell'essere umano è generalmente declinata in doveri di sorveglianza e di eventuale intervento, quando necessario. La tipologia di attività è decisiva per determinare come tali doveri si sostanzino in concreto, in particolare in riferimento all'importanza della funzione automatizzata, ai rischi connessi e al costo di eventuali malfunzionamenti. Infatti, forme il controllo rigoroso da parte di esseri umani sono generalmente presenti in contesti percepiti, per ragioni differenti, come particolarmente delicati, dei quali possono essere validi esempi i cicli industriali almeno parzialmente automatizzati o i sistemi di pilotaggio automatico usati nell'aviazione civile[712]. In numerosi casi, inoltre, l'intervento umano, e la sorveglianza che implica a monte, risulta necessario per perfezionare o portare a termine le azioni di sistemi automatizzati che, allo stato dell'arte, non possono prescindere dalla partecipazione di un operatore in carne ed ossa per portare a termine i loro compiti[713].

D'altro canto, non si può trascurare che proprio la totale automazione di determinate procedure, e, di conseguenza, la liberazione di tempo ed energie degli esseri umani che ne consegue, è stata ed è

[711] Svolgono considerazioni simili C. RUDIN, Stop *explaining black box machine learning models for high stakes decisions and use interpretable models instead*, in *Nature Machine Intelligence*, 1, 5, 2019, p. 206-215; A.D. SELBST, J. POWLES, *Meaningful information and the right to explanation cit.*; L. EDWARDS, M. VEALE, *Enslaving the algorithm: from a "right to an explanation" to a "right to better decisions"? cit.*

[712] Cfr ad esempio A. OSUNWUSI, *Aviation Automation and CNS/ATM-related Human-Technology Interface: ATSEP Competency Considerations*, in *International Journal of Aviation, Aeronautics, and Aerospace*, 6, 4, 2019, https://commons.erau.edu/ijaaa/vol6/iss4/13 (11 agosto 2022); Y. CHENG ET AL., *Reliability Prediction and Safety Evaluation of ATC Automation System*, in *2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, 2020, https://ieeexplore.ieee.org/abstract/document/9368710 (11 agosto 2022).

[713] Un tipico esempio del coinvolgimento dell'essere umano in attività parzialmente automatizzate è rappresentato dalle moderne catene di montaggio robotizzate, cfr. I. KURNIAWAN, J. M. POWER, V. I. CANECA, T. D. SOFIANTI, *System Modelling and Simulation for Study of Human-Machine Collaboration Technologies Implementation on Assembly Line*, in *Proceedings of the International Conference on Engineering and Information Technology for Sustainable Industry*, Association for Computing Machinery, New York, 2020, https://doi.org/10.1145/3429789.3429799 (11 agosto 2022).

alla base dello sviluppo tecnico-industriale degli ultimi due secoli[714]. Una vastissima gamma di attività è svolta interamente da macchine in completa assenza di controllo da parte dell'essere umano, o con forme di sorveglianza e verifica solamente episodica, consistenti, nella pratica, in attività di semplice manutenzione. Ciò non suscita particolare allarme, e, anzi, lo sviluppo di tecnologie di tal genere è spesso concepito come una desiderabile evoluzione economica e sociale, per l'emancipazione dell'uomo da lavori pesanti e pericolosi che ne deriva. Le voci critiche sul tema si concentrano sui possibili effetti a medio e lungo termine sul piano occupazionale, teorizzando una diminuzione del numero complessivo dei posti di lavoro, o comunque un mutamento del mercato di questi ultimi che finerebbe per pregiudicare i soggetti meno qualificati[715]. L'automazione di tali attività, però, non appare preoccupante: un sistema di irrigazione non è certo percepito come un rischio, e lo stesso vale per un cancello automatizzato, uno sportello bancomat, e numerosissimi altri casi della vita quotidiana, in cui l'intervento umano è solo eventuale, e spesso dev'essere sollecitato dall'utente in caso di malfunzionamenti.

L'avvento dell'intelligenza artificiale ha, ovviamente, un effetto dirompente su questo assetto dell'automazione, definibile di tendenziale accettazione. Infatti, le attività che è possibile automatizzare, in tutto o in parte, con l'ausilio dell'intelligenza artificiale appaiono subito qualitativamente differenti, perché riguardano compiti comunemente associati alla creatività, alle competenze e alle abilità di valutazione e giudizio dell'essere umano. Non si tratta, dunque, del semplice svolgimento da parte delle macchine diprocedimenti che, per quanto complessi, risultano replicabili col solo lavoro manuale, ma di compiti per i quali è sempre stato necessario l'impiego, a vari livelli, dell'intelligenza umana. La circostanza solleva numerosi interrogativi, ad esempio sul tema, già citato, del possibile impatto sul mercato del lavoro. Tra i temi di discussione più rilevanti, in primo luogo dal punto di vista giuridico, rientrano, come subito si dirà, le eventuali forme di controllo, sorveglianza e intervento umano sul funzionamento di tali tecnologie[716]. Le innovazioni connesse all'intelligenza artificiale, infatti, impongono un ripensamento del ruolo assegnato

[714] Cfr. ad esempio F. N. DAVID, *Forces of Production: A Social History of Industrial Automation*, New York, 2017.

[715] Cfr. ad es. M. FORD, *Rise of the Robots: Technology and the Threat of a Jobless Future*, New York, 2015; D. M. WEST, *What happens if robots take the jobs? The impact of emerging technologies on employment and public policy*, in *Center of Technology Innovation at Brookings*, Oct. 2015, https://bit.ly/3Gue9yS (20ottobre 2021); D. H. AUTOR, *Why are there still so many jobs? The history and future of workplace automation*, in *Journal of Economic Perspectives*, 29, 2015, p. 3 ss.

[716] Si vedano M. L. JONES, *The right to a human in the loop: Political constructions of computer automation and personhood*, in *Social Studies of Science*, 47, 2, 2017, p. 216-239; C.CASONATO, *AI and constitutionalism: the challenges ahead*, in B. BRAUNSCHWEIG, M. GHALLAB, *Reflections on Artificial Intelligence for Humanity*, Berlino, 2021, p. 127-149;A. ODENNINO, *Decisioni algoritmiche e prospettive internazionali di valorizzazione dell'intervento umano*, in *DPCE-online*, 1, 2020, p. 199-217; S. AMATO, *Biodiritto 4.0*, Torino, 2020, p. 81 ss.; A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal - Rivista di BioDiritto*, 1, 2019.

all'essere umano rispetto all'automazione, per l'inedito livello di autonomia delle macchine che implicano.

Nella letteratura scientifica d'ambito informatico e ingegneristico è comune riferirsi al coinvolgimento dell'essere umano in un processo automatizzato con l'espressione *human in the loop*[717]. La partecipazione dell'umano nel ciclo di funzionamento di una tecnologia è spesso considerata come un obiettivo di progettazione, in grado di avere un impatto positivo sui risultati ottenibili con quest'ultima. Studi recenti, ad esempio, hanno evidenziato il potenziale rappresentato dall'essere umano, in termini di dati utili all'elaborazione, per lo sviluppo di sistemi di *machine learning* e di tecnologie riconducibili, in generale, all'Internet of Things (da qualcuno appropriatamente definito, in questa accezione, *Internet of All*)[718]. In quest'ottica, l'essere umano non andrebbe concepito come semplice utente passivo di tali tecnologie, ma andrebbe attivamente coinvolto nel loro sviluppo con forme di progettazione che favoriscano l'interazione continua con esse, onde favorire la raccolta di dati per il loro continuo miglioramento. La persona fisica, dunque, dovrebbe concepirsi come elemento a pieno titolo di un sistema tecnologico integrato. Un'evoluzione simile a quanto già avvenuto, in parte, nell'ambiente *online* con la diffusione dell'internet interattivo, che ha di certo favorito lo sviluppo di applicazioni tecnologiche di inedita efficacia e indubbia utilità, ma con possibili conseguenze, in assenza di una regolazione consapevole, per una vasta gamma di diritti fondamentali, come già analizzato nella seconda parte di questo lavoro[719].

Un'altra concezione dello *human in the loop* rinvenibile nella letteratura tecnico-scientifica, invece, lo concepisce come un elemento di rischio da razionalizzare per il funzionamento di determinati sistemi. È una visione riscontrabile, in particolare, in materia di tecnologie impiegate con funzioni di sicurezza, non necessariamente basate sull'intelligenza artificiale, come sistemi che regolano l'accesso ad aree riservate o tecniche crittografiche per la protezione della riservatezza di determinati documenti associate a password. L'umano è comunemente percepito come "l'anello debole" di qualunque procedura di sicurezza, e l'assunto è giustificato dalla circostanza che l'errore umano è di gran lunga la causa più comune di ogni violazione o malfunzionamento di tali sistemi.

[717] Cfr. ad esempio X. WU, L. XIAO, Y. SUN, J. ZHANG, T. MA, L. HE, *A survey of human-in-the-loop for machine learning*, in *Future Generation Computer Systems*, 135, 2022, p. 364-381; F.M. ZANZOTTO,*Viewpoint: Human-in-the-loop Artificial Intelligence*, in *Journal of Artificial Intelligence Research*, 64, 2019, p. 243-252; W. LI, D. SADIGH, S.S. SASTRY, S.A. SESHIA, *Synthesis for Human-in-the-Loop Control Systems*, in E. ÁBRAHÁM, K. HAVELUND (A CURA DI), *Tools and Algorithms for the Construction and Analysis of Systems*, Berlino-Heidelberg, 2014, p. 470-484; M. GIL, M. ALBERT, J. FONS, V. PELECHANO, *Engineering human-in-the-loop interactions in cyber-physical systems*, in *Information and Software Technology*, 126, 2020.

[718] Cfr. D.S. NUNES, P. ZHANG, J. SÁ SILVA, *A Survey on Human-in-the-Loop Applications Towards an Internet of All*, in *IEEE Communications Surveys & Tutorials*, 17, 2, 2015, p. 944-965.

[719] V. *supra* p. 75 ss. Cfr. inoltre i citati D. DINUCCI, *Fragmented future cit.*; P. MAGRASSI, T. BERG, *A World of Smart Objects cit.*; K. ASHTON, *That "Internet of Things" thing cit.*; D. UCKELMANN, M. HARRISON, F. MICHAHELLES (A CURA DI), *Architecting the Internet of Things cit*.

La totale automazione, così, è spesso identificata in modo esplicito come l'obiettivo tendenziale di un sistema di sicurezza ottimale, e, per quanto riguarda i contesti in cui ciò non sia possibile, esiste un intero filone di studi dedicato alla progettazione di sistemi in cui l'integrazione del necessario elemento umano si svolga con modalità che limitino al minimo il pericolo ad essa connesso[720].

La produzione scientifica d'ambito tecnologico, dunque, concepisce lo *human in the loop*, a seconda dei casi, come un'obiettivo da raggiungere per lo sviluppo di sistemi più integrati ed efficienti o come un male necessario del quale circoscrivere le possibili conseguenze negative. L'idea che esso, invece, possa rappresentare una garanzia verso forme di delega troppo ampia alla tecnologia, con effetti potenzialmente disumanizzanti, hatrovato spazio solo di recente, e principalmente in studi d'ambito filosofico, etico e giuridico (in quest'ultima accezione, come vedremo, è diffusa anche l'espressione *human oversight*)[721]. L'avvento dell'intelligenza artificiale ha rappresentato la principale spinta propulsiva di tali riflessioni. Le ragioni sono intuitive e sonostate tutte già menzionate, da altri punti di vista, nel corso del lavoro: la possibilità di automazione di un crescente numero di attività in precedenza esclusivamente intellettuali, ad esempio, ha portato all'impiego di tecnologie totalmente o parzialmente autonome in contesti decisionali percepiti come discrezionali;la base di funzionamento essenzialmente statistica di molti sistemi di apprendimento automaticofa apparire qualitativamente diverso, rispetto a tecnologie automatizzate più risalenti e riconducibili ad altri approcci, il loro coinvolgimento in ambiti intrinsecamente pericolosi, come l'industria pesante;lo sviluppo di tecnologie di intelligenza artificiale sempre nuove, destinate a interagire col pubblico, nell'ambiente online o nel mondo fisico, fa sembrare prossimo un mondo in cui l'automazione abbia un peso sempre crescente nella vita dell'individuo. La riflessione etico-giuridica sul controllo residuo in capo all'essere umano nasce dalla presa di coscienza delle possibili conseguenze negative di questi mutamenti tecnologici, anch'esse, in larga parte, già analizzate nelle parti precedenti di questo studio: i sistemi automatizzati possono sempre presentare guasti e malfunzionamenti; molte tecnologie intelligenti presentano un intrinseco margine d'errore, pur di estensione ridotta; l'analisi dei dati condotta con strumenti di intelligenza artificiale può essere viziata da *bias* di varia natura. Alcune ricerche,

[720] Cfr. ad esempio L.F. CRANOR, *A framework for reasoning about the human in the loop*, in *Proceedings of the 1st Conference on Usability, Psychology, and Security*, USENIX Association, USA, 2008, https://dl.acm.org/doi/10.5555/1387649.1387650 (10 agosto 2022). Per una critica di questa prospettiva cfr. A. PAWLICKA, M. PAWLICKI, R. KOZIK, M. CHORAŚ, *Human-driven and human-centred cybersecurity: policy-making implications»,* in *Transforming Government: People, Process and Policy*, 2022 (in corso di pubblicazione), disponibile in: https://doi.org/10.1108/TG-05-2022-0073 (10 agosto 2022).

[721] L'espressione, come si dirà, è stata adottata anche dalla Proposta di Regolamento in materia di intelligenza artificiale presentata dalla Commissione Europea il 21 aprile 2021, il cui art. 14, nella versione di lingua inglese, è appunto rubricato *Human oversight*. In letteratura si rinvia, *ex multis*, ai già menzionati M. L. JONES, *The right to a human in the loop cit.*; C.CASONATO, *AI and constitutionalism: the challenges ahead cit.*; A. ODENNINO, *Decisioni algoritmiche e prospettive internazionali di valorizzazione dell'intervento umano cit.*; A. SIMONCINI, *L'algoritmo incostituzionale cit.*

inoltre, hanno evidenziato come l'automazione diffusa potrebbe avere una conseguenza particolarmente infida: la graduale perdita, da parte dell'essere umano, della capacità di svolgere le attività delegate alla macchina, magari tradizionalmente legate a formazioni specialistiche d'alto livello (c.d. *deskilling*) [722]. Operatori umani che in precedenza svolgevano tali mansioni frequentemente vedrebbero dimunire le occasioni in cui farlo, dimenticando a poco a poco le nozioni necessarie; i loro colleghi più giovani avrebbero molte meno occasioni di mettere in pratica le competenze teoriche aquisite in tali ambiti; si giungerebbe, sul medio-lungo periodo, alla scomparsa del patrimonio di conoscenze connesso a tali attività dal percorso formativo in questione. Le eventuali conseguenze del fenomeno sono intuitive: risulterebbe sempre più difficile sostituire il sistema automatizzato in caso di incidenti o malfunzionamenti e potrebbe risultare complessa anche la semplice individuazione di questi ultimi, per la crescente rarità delle competenze necessarie.

4.2. I distinti livelli di controllo umano sul sistema e il legame con la sua spiegabilità

In seno al dibattito sul possibile impatto etico, sociale ed economico delle tecnologie intelligenti, dunque, il mantenimento di una soglia minima di controllo, sorveglianza, ed eventuale possibilità d'intervento dell'essere umano è concepito come un presidio nei confronti dei possibili effetti avversi delle nuove forme di automazione, potenzialmente lesivi di una vasta gamma di diritti fondamentali[723]. Il requisito dello *human in the loop* dovrebbe mettere al riparo dalle conseguenze di eventuali difetti di funzionamento, di qualunque origine e natura, e permettere, più in generale, la conservazione di un livello minimo di dominio dell'essere umano sulla realtà che lo circonda. Le modalità concrete con cui predisporre questo genere di garanzia antropocentrica sono, però, destinate a variare a seconda della tecnologia coinvolta, e non paiono, in generale, affatto semplici. Non tutte le applicazioni dell'intelligenza artificiale, infatti, paiono potersi sottoporre con facilità al controllo dell'essere umano, in ragione della loro complessità tecnica – in virtù della quale lo *human in the loop* potrebbe necessitare di competenze specialistiche non comuni – o della loro mancanza di trasparenza, che potrebbe renderle scarsamente comprensibili, nell'accezione vista ai paragrafi precedenti. Come si dirà, infatti, tra la possibilità di ricavare una spiegazione del comportamento di un sistema e la sottoponibilità di quest'ultimo a un determinato livello di

[722] Cfr. ad esempio J. LU, *Will Medical Technology Deskill Doctors?*, in *International Education Studies*, 9, 7, 2016, p. 130–134; J. LEVY, A. JOTKOWITZ, I. CHOWERS, *Deskilling in Ophthalmology Is the Inevitable Controllable?*, in *Eye*, 33, 3, 2019, p. 347–348; S. DE PAOLI, *Automatic-Play and Player Deskilling*, in *MMORPGs,Game Studies*, 13, 1, 2013; E. SINAGRA, F. ROSSI, D. RAIMONDO, *Use of Artificial Intelligence in Endoscopic Training: Is Deskilling a Real Fear?*, in *Gastroenterology* 160, 6, 2021, p. 2212 ss.Cfr. anche *supra*, p. 144-145.

[723] Cfr. ancora, ad esempio, M. L. JONES, *The right to a human in the loop cit.*; C.CASONATO, *AI and constitutionalism: the challenges ahead cit.*; A. ODENNINO, *Decisioni algoritmiche e prospettive internazionali di valorizzazione dell'intervento umano cit.*; A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà cit.*; C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni cit.* e *Costituzione e intelligenza artificiale cit*.

controllo da parte dell'essere umano esiste un ovvio legame. Inoltre, la predisposizione delle condizioni necessarie perché una determinata tecnologia, a prescindere dal filone dell'intelligenza artificiale cui essa appartenga, sia soggetta alla sorveglianza di un efficace *human in the loop* potrebbe risultare particolarmente complicata per la problematica, già menzionata, della c.d. distorsione dell'automazione[724]. I risultati di alcuni studi fanno ipotizzare che un essere umano chiamato a utilizzare sistemi avanzati in un'attività della quale è responsabile tenda a non mettere in discussione le indicazioni o il comportamento di questi ultimi, anche in situazioni in cui risultino errati e sembri ragionevole aspettarsi che, senza la tecnologia, l'operatore non avrebbe commesso l'errore, o si sarebbe avveduto della sua presenza[725]. La possibilità concreta di sottoporre l'intelligenza artificiale a un effettivo controllo umano, dunque, pare complicata da questo fenomeno psicologico, di fronte al quale sembra necessario lo sviluppo di strategie di informazione specifica, che aumentino la consapevolezza della possibilità di scivolare in questa forma di eccessivo affidamento alla tecnologia. Si tratta, come si dirà più approfonditamente al paragrafo successivo, di un'ipotesi che anche il Legislatore europeo sembra in procinto di prendere in considerazione[726].

La varietà e complessità delle possibili modalità di interazione tra essere umano e macchina rendono ipotizzabili varie forme di controllo umano, che è possibile distinguere in base all'intensità dei poteri dello *human in the loop* sul sistema. In particolare, può ipotizzarsi una suddivisione in quattro macrolivelli delle modalità di sorveglianza umana sull'intelligenza artificiale, prendendo spunto dauna distinzione elaborata nelle *Linee guida etiche* del 2018 del Gruppo di Esperti di Alto

[724] Cfr. in primo luogo quanto esposto *supra*, p. 143 ss. e n. 486. La dicitura "distorsione dell'automazione" è impiegata, come subito si dirà (cfr. n. 723), dalla Proposta di Regolamento in materia di intelligenza artificiale. Il magistrato e studioso di diritto francese Antoine Garapon, invece, si riferisce al fenomeno con l'evocativa espressione "effet moutonnier", traducibile in italiano come *effetto pecorone*. In lingua inglese, invece, il problema è talvolta descritto come *over-reliance*. Cfr. J. WU; J. THORNE-LARGE; P. ZHANG, *Safety first: The risk of over-reliance on technology in navigation*, in *Journal of Transportation Safety & Security*, 14, 7, 2022; A. GARAPON, J. LASSEGUE, *Justice Digitale cit.*; E. FRONZA, *"Code is Law" cit.*; J. DE CODT, *Justice et algorithme cit.*; B. MARCHETTI, *La garanzia dello human in the loop cit.*; C. CASONATO, *L'intelligenza artificiale e il diritto pubblico comparato ed europeo cit.*; C. CASONATO, *Giustizia e intelligenza artificiale: considerazioni introduttive cit.*

[725] Cfr. ad esempio J. WU; J. THORNE-LARGE; P. ZHANG, *Safety first: The risk of over-reliance on technology in navigation cit.*; M. R. ENDSLEY, *Automation and Situation Awareness*, in R. PARASURAMAN, M. MOULOUA, *Automation and Human Performance: Theory and Applications*, Boca Raton (US), 1996. Per risultati recenti che ipotizzano che il *bias* sia più evidente quando l'indicazione dell'algoritmo confermi pregiudizi esistenti nel decisore umano, e possa, invece, ridursi fino a scomparire quando ciò non accada o il soggetto sia consapevole della possible effetto distorsivo dell'algoritmo, cfr. S.A. BARKAT, M. BUSIOC, *Human–AI Interactions in Public Sector Decision Making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice*, in *Journal of public administration research and theory*, 2022, https://doi.org/10.1093/jopart/muac007 (12 ottobre 2022).

[726] Infatti, l'art. 14 della Proposta, come già detto rubricato *human oversight*, tra i molti requisiti che impone per i sistemi ad alto rischio prevede che l'utente sia messo in condizione di (par. 4 lett. b): «restare consapevole della possibile tendenza a fare automaticamente affidamento o a fare eccessivo affidamento sull'output prodotto da un sistema di IA ad alto rischio ("distorsione dell'automazione"), in particolare per i sistemi di IA ad alto rischio utilizzati per fornire informazioni o raccomandazioni per le decisioni che devono essere prese da persone fisiche», cfr. *infra* p. 224.

Livello sull'IA nominato dalla Commissione Europea[727]. In primo luogo, uno scenario di *human out of the loop*, corrispondentealla totale automazione, in cui ogni possibilità di sorveglianza ed intervento umano, anche minimo, sia assente. In secondo luogo, un livello di controllo definibile *human on the loop*, in cui l'essere umano, una volta esaurito il ciclo di progettazione del sistema – al quale, ovviamente, partecipa attivamente – si limita a monitorarne il corretto funzionamento e, se necessario, a interromperlo in tutto o in parte. In terzo luogo, l'approccio che le menzionate *Linee guida* definiscono *human in the loop*, nonostante, come già detto, il termine sia comunemente impiegato, nella letteratura specialistica sia di lingua inglese che di lingua italiana, per indicare il tema del controllo umano sulla tecnologia in senso lato. Questo terzo livello di controllo prevede che l'essere umano abbia la possibilità di intervenire in ogni ciclo di funzionamento del sistema, anche modificandone i parametri forniti come input, e, di conseguenza, risultati e comportamento. Infine, una situazione di *human in command*, in cui l'automazione del sistema è limitata alla possibilità di portare a termine in autonomia, una volta avviata e qualora non vi siano successive interferenze da parte dell'essere umano, l'attività per cui è stato progettato. L'intervento dello *human in command* è richiesto per l'inizio di ogni sessione di funzionamento e la possibilità di controllo rimane totale in ogni fase: è l'umano a decidere se, come, e quando utilizzare la tecnologia, valutandone di volta in volta i possibili effetti.

Questa categorizzazione fa apparire immediatamente evidente il legame con il tema della trasparenza e comprensibilità dei sistemi di intelligenza artificiale. I livelli più alti di controllo, infatti, sembrano applicabili solamente di fronte ad applicazioni tecnologiche non opache. In particolare, l'interpretabilità del sistema – intesa come possibilità di conoscerne gli stati interni e ricavare da questa consapevolezza informazioni sui suoi *output* – sembra decisiva per permettere all'essere umano di intervenire sul ciclo di funzionamento, modificandone le modalità al fine di influenzarne il risultato finale. L'eventuale spiegabilità di tale risultato con strategie di *explainable artificial intelligence* in tutto o in parte *model-agnostic* non pare influire di molto sulla circostanza, potendo garantire informazioni di certo utili al controllo passivo dell'essere umano sul funzionamento del sistema, ma da cui non potrebbero ricavarsi agevolmente elementi per interferire su quest'ultimo. Interi sottosettori dell'intelligenza artificiale, dunque, in primo luogo il filone delle reti neurali profonde, non sembrano compatibili coi livelli di *human oversight* in precedenza identificati con le espressioni *human in the loop* e *human in command*, e paiono sottoponibili unicamente a forme di monitoraggio sul loro funzionamento del tipo definito *human on the loop*. Preme evidenziare che quest'ultima garanzia non appare da sottovalutare e, anzi, potrebbe definirsi il nucleo essenziale dello *human oversight*: controllo e sorveglianza umana mantengono il loro

[727] Cfr. HIGH LEVEL EXPERT GROUP ON AI, *Ethics Guidelines for Trustworthy Artificial Intelligence cit.*, p. 18.

significato solamente se consistono, almeno, nella possibilità di interrompere il funzionamento del sistema quando appaia anomalo, errato, o rischioso, e di ignorare, discostarsi in parte o sovvertire radicalmente i suoi risultati.

Le concrete modalità di controllo umano sul sistema potrebbero, chiaramente, variare di molto a seconda dell'applicazione tecnologica coinvolta, e dovrebbe essere compito del diritto, come si vedrà più approfonditamente al paragrafo successivo, predisporre un assetto normativo idoneo ad assicurare, al contempo, flessibilità ed effettività della tutela. Il requisito dello *human in the loop*, dunque, dovrebbe prima di tutto rappresentare l'argine etico-giuridico verso forme di delega totale e irreversibile alla macchina – come nel caso definito, nella tassonomia sopra ipotizzata, *human out of the loop* – rischiose per una vasta gamma di diritti fondamentali. Da questo punto di vista, l'eventuale livello di opacità della tecnologia gioca, come già detto, un ruolo essenziale, poiché determina l'intensità dei poteri sul sistema configurabili in capo all'essere umano. Si deve sottolineare, in ogni caso, che le categorie della spiegabilità e del controllo umanonon sono di certo sovrapponibili: un livello minimo di sorveglianza, intesa almeno come possibilità di interromperne il funzionamento, è ipotizzabile anche su tecnologie *black-box* di fronte alle quali, per ipotesi, i risultati delle tecniche di *explainable artificiale intelligence* apparissero particolarmente insoddisfacenti. Questa basilare garanzia di uno "*stop button*", anzi, potrebbe rappresentare un presidio elementare e irrinunciabile in tutti i contesti in cui spiegabilità e interpretabilità del sistema appaiano particolarmente ridotte, ma il suo utilizzo sembri, per i valori concretamente in gioco, comunque auspicabile, portando, in esito al bilanciamento, al parziale sacrificio del già analizzato diritto alla spiegazione[728]. Proprio alla possibilità di un'eventuale giuridicizzazione, al livello di diritto fondamentale, del requisito dello *human on the loop* o *human oversight*, e dunque alla sua idoneità a dialogare con interessi di pari rango, sarà dedicato il prossimo paragrafo.

4.3. Il parziale riconoscimento nel diritto positivo del diritto al controllo umano sul sistema

Allo stato dell'arte, i principali esempi di giuridicizzazione del requisito tecnico dello *human oversight* si rinvengono, ancora una volta, nel GDPR e nella *Directive on automated decision-*

[728]La necessità di predisporre un pulsante d'arresto come garanzia minima di controllo umano è prevista anche dalla *checklist* che conclude le menzionate HIGH LEVEL EXPERT GROUP ON AI, *Ethics Guidelines for Trustworthy Artificial Intelligence cit.*, p. 31. In letteratura evidenziano l'importanza del requisito, tra gli altri, D. HADFIELD-MENELL, A. DRAGAN, P. ABBEEL, S. RUSSELL, *The Off-Switch Game*, https://arxiv.org/abs/1611.08219v3, 2016 (11 agosto 2022); B. SHNEIDERMAN, *Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy*, in *International Journal of Human–Computer Interaction*, 36, 6, 2020; L. ORSEAU, M. ARMSTRONG, *Safely interruptible agents*, in *Conference on Uncertainty in Artificial Intelligence*, 2016, https://bit.ly/3eDFDbV (11 agosto 2022). Per una visione più critica, che mette in guardia dal rischio di considerare ogni problema relativo al controllo umano sui sistemi intelligenti risolto con la semplice predisposizione di uno *stop-button*, cfr. T. ARNOLD M. SCHEUTZ, *The "big red button" is too late: an alternative model for the ethical evaluation of AI systems*, in *Ethics and Information Technology*, 20, 1, 2018, https://link.springer.com/article/10.1007/s10676-018-9447-7 (11 agosto 2022).

making canadese. Per quanto riguarda il GDPR viene in gioco, in particolare, il già più volte menzionato art. 22, che, come già visto, sancisce il divieto, in linea di principio, di sottoporre l'individuo a un trattamento dei suoi dati personali che consista in una decisione automatizzata con «effetti giuridici che lo riguardano, o incida in modo analogo significativamente sulla sua persona»[729]. Come già riportato, il secondo paragrafo della norma introduce rilevantissime eccezioni alla previsione, stabilendo che la decisione automatizzata sia permessa per finalità di conclusione o esecuzione di un contratto tra titolare e interessato; sia resa lecita dal diritto dell'Unione o degli stati membri; si fondi sul consenso esplicito dell'interessato[730]. Per due di questi casi, la conclusione o esecuzione di un accordo contrattuale e il trattamento basato sul consenso, il par. 3 dello stesso art. 22 prevede, tra le misure a tuteladei «diritti, le libertà e i legittimi interessi dell'interessato» che devono accompagnare la decisione automatizzata, «almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione»[731].

Il fatto che la possibilità di richiedere l'intervento umano sia garantita anche nei casi in cui la decisione automatizzata è eccezionalmente ammessa, porta a pensare che il primo paragrafo dell'art. 22, nello statuire il diritto dell'interessato a non essere sottoposto a una decisione basata sul trattamento automatizzato, imponga di assicurare un controllo umano effettivo e non meramente potenziale. Esso potrà assumere, come visto al paragrafo precedente, forme differenti a seconda delle caratteristiche della tecnologia coinvolta, e in particolare della sua eventuale opacità.

L'art. 22 par. 3, inoltre, riconosce all'interessato il diritto di esprimere la propria opinione e di contestare la decisione che lo riguarda. Anche quest'ultima posizione giuridica pare doversi ricondurre, in ultima analisi, all'area del controllo umano sul sistema, interpretandola come possibilità di sottoporre la decisione a una revisione umana a cura dello stesso titolare del trattamento[732]. L'attribuzione di un diverso significato, infatti, renderebbe superflua la norma. Un

[729] Per il testo completo dell'art. 22 GDPR cfr. *supra*, p. 163, n. 543 e p. 202, n. 659.

[730] Cfr. ancora *supra*, in materia di diritto alla spiegazione, p. 185 ss. Come già riportato, considerano eccessiva l'ampiezza di tali eccezioni C. CASONATO, *Costituzione e intelligenza artificiale cit.*, p. 723-724; A. SIMONCINI, *L'algoritmo incostituzionale cit.*, p. 79 ss.

[731] Per chiarezza, si riporta nuovamente l'art. 22 par. 3 GDPR: «Nei casi di cui al paragrafo 2, lettere a) e c), il titolare del trattamento attua misure appropriate per tutelare i diritti, le libertà e i legittimi interessi dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione». Per dei commenti si rimanda nuovamente a A. ODDENINO, *Decisioni algoritmiche e prospettive internazionali di valorizzazione dell'intervento umano cit.*; A. CAIA, *Art. 22 cit.* Può essere significativo notare che la parte dell'art. 22 par. 3 citata compare, con identica formulazione, anche nel Cons. 71 GDPR, meno che per l'elemento del «diritto a ottenere una spiegazione della decisione», che, come evidenziato trattando tale posizione giuridica, è stato escluso dall'articolato del Regolamento.

[732] Il legame tra l'elemento dell'intervento umano e la possibilità di contestare la decisione, e gli oneri che quest'ultima pone in capo al titolare del trattamento, sono rilevati anche dalle ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 cit.*, p. 27: «Human intervention is a key element. Any review must be carried out by someone who has the appropriate authority

riesame in tutto o in parte automatizzato svuoterebbe di senso la disposizione, la cui *ratio* è, appunto, la tutela da eventuali conseguenze infauste dell'automazione; non ritenere il titolare del trattamento onerato della revisione della decisione priverebbe di contenuto il diritto a contestare quest'ultima, che si risolverebbe nella semplice ricognizione del diritto ad agire in ogni sede opportuna a tutela dei propri interessi, già garantito, ovviamente, dal diritto europeo e dagli ordinamenti interni degli stati membri.

Come già analizzato nei precedenti paragrafi, per quanto riguarda i trattamenti consistenti in decisioni automatizzate che fondino la loro liceità sul diritto dell'Unione e degli stati membri, l'art. 22 par. 2 GDPR non definisce, nemmeno nei tratti essenziali, la natura delle garanzie a tutela di diritti, libertà e interessi legittimi che impone di predisporre. Solo una minoranza dei paesi membri, finora, ha legiferato sulla base di questa disposizione, prevedendo la possibilità di decisioni automatizzate attraverso l'elaborazione di dati personali in circostanze ulteriori rispetto a quelle stabilite dal Regolamento europeo[733]. Nella quasi totalità dei casi, tali inziative di diritto interno sono state accompagnate dall'introduzione di cautele assimilabili, per quanto riguarda controllo, sorveglianza e intervento umani, a quanto stabilito dalle norme dell'Unione, con la parziale eccezione della Francia, che, nelle proprie norme interne, pare avere almeno in parte trascurato l'elemento dello *human oversight*[734]. Il riferimento, in particolare, è al già visto art. 47 della *Loi n. 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés*, novellato dalla *Loi n. 2018-493 du 20 juin 2018 relative à la protection des données personnelles*, di adeguamento al GDPR[735]. Come già detto, il primo comma della disposizione chiude la porta a ogni ipotesi di automazione della decisione giudiziaria sul comportamento di una persona fisica; la prima parte del

---

and capability to change the decision. The reviewer should undertake a thorough assessment of all the relevant data, including any additional information provided by the data subject»,

[733]Cfr. ancora G. MAUGERI, *Automated decision-making in the EU Member States: the right to explanation and other "suitable safeguards" in the national legislations cit.*

[734]*Ibidem*, p. 12 ss.

[735]Per chiarezza, si riporta nuovamente il testo dell'art. 47 della *Loi informatique* attualmente in vigore: «1. Aucune décision de justice impliquant une appréciation sur le comportement d'une personne ne peut avoir pour fondement un traitement automatisé de données à caractère personnel destiné à évaluer certains aspects de la personnalité de cette personne. 2. Aucune décision produisant des effets juridiques à l'égard d'une personne ou l'affectant de manière significative ne peut être prise sur le seul fondement d'un traitement automatisé de données à caractère personnel, y compris le profilage, à l'exception : 1° Des cas mentionnés aux a et c du 2 de l'article 22 du règlement (UE) 2016/679 du 27 avril 2016, sous les réserves mentionnées au 3 du même article 22 et à condition que les règles définissant le traitement ainsi que les principales caractéristiques de sa mise en œuvre soient communiquées, à l'exception des secrets protégés par la loi, par le responsable de traitement à l'intéressé s'il en fait la demande; 2° Des décisions administratives individuelles prises dans le respect de l'article L. 311-3-1 et du chapitre Ier du titre Ier du livre IV du code des relations entre le public et l'administration, à condition que le traitement ne porte pas sur des données mentionnées au I de l'article 6 de la présente loi. Ces décisions comportent, à peine de nullité, la mention explicite prévue à l'article L. 311-3-1 du code des relations entre le public et l'administration. Pour ces décisions, le responsable de traitement s'assure de la maîtrise du traitement algorithmique et de ses évolutions afin de pouvoir expliquer, en détail et sous une forme intelligible, à la personne concernée la manière dont le traitement a été mis en œuvre à son égard. 3. Par dérogation au 2° du présent article, aucune décision par laquelle l'administration se prononce sur un recours administratif mentionné au titre Ier du livre IV du code des relations entre le public et l'administration ne peut être prise sur le seul fondement d'un traitement automatisé de données à caractère personnel».

secondo comma ribadisce e specifica le condizioni di liceità previste dall'art. 22 par. 2 lett. a) e c) del GDPR, inserendo oneri informativi supplementari in capo al titolare; la seconda parte funge da base di liceità per decisioni amministrative totalmente automatizzate, a patto che il trattamento non coinvolga dati c.d. particolari e rispetti gli stringenti requisiti volti ad assicurare una spiegazione all'utente e un livello accettabile di comprensibilità della tecnologia coinvolta previsti dagli artt. L. 311-3-1 ss. del *Code des relations entre le public et l'administration*, come riformato dalla *Loi n. 2016-1321 du 7 octobre 2016 pour une République numérique*[736]. In quest'ultimo caso – l'unico a estendere l'area di legittimità della decisione automatizzata rispetto al Regolamento europeo – la norma non esplicita alcuna garanzia di controllo e intervento umano sulla decisione, nemmeno su richiesta dell'interessato. L'unica parte del testo legislativo che valorizzi l'elemento umano, infatti, attiene, come già riportato, alle competenze specialistiche che il funzionario responsabile del procedimento dovrebbe avere – non è chiaro in base a quale formazione – al fine di: «pouvoir expliquer, en détail et sous une forme intelligible, à la personne concernée la manière dont le traitement a été mis en œuvre à son égard». L'ordinamento francese, dunque, affianca a una tutela del diritto alla spiegazione in materia di decisione amministrativa automatizzata particolarmente forte, garanzie particolarmente deboli riguardo allo *human oversight*[737], non prevedendo, in particolare, il diritto dell'interessato a richiedere l'intervento di un essere umano nella decisione, a differenza di quanto previsto, nel rispettivo ambito di applicazione, dall'art. 22 par. 3 del GDPR. L'opportunità di una revisione umana della decisione pare rinviata alla sempre presente possibilità di ricorso amministrativo avverso di essa. Da questo punto di vista, è significativo che l'ultimo comma dell'art. 47 della *Loi informatique et libertés* in esame escluda, in deroga al comma precedente, che un ricorso amministrativo possa mai essere deciso «sur le seul fondement d'un traitement automatisé de données à caractère personnel».

Volgendo lo sguardo all'ordinamento canadese, la *Directive on automated decision-making* impone, al paragrafo 6.3.9, di «ensuring that an Automated Decision System allows for human intervention, when appropriate, as per Appendix C»[738]. Il citato Appendix C precisa i confini di tale *human intervention*, in funzione dei livelli di rischio associati al sistema definiti, come già visto, dal precedente Appendix B[739]. Per le due classi di rischio più basse – ovvero quando l'impatto sui

[736] Anche in questo caso, per il testo completo di tali norme si rimanda all'analisi in materia di diritto alla spiegazione già svolta *supra*, p. 205 ss.

[737] L'indifferenza dell'ordinamento francese al requisito del controllo umano, infatti, risulta in antitesi a quanto disposto in materia di diritto alla spiegazione, del quale tale sistema, attraverso le proprie norme di diritto interno, ha predisposto, come già visto, una tutela almeno in parte rafforzata.

[738] Il paragrafo riportato, significativamente, apre la sezione della Direttiva rubricata *Ensuring human intervention.*

[739] Per i dettagli della classificazione dei sistemi di decisione automatizzata in quattro livelli di rischio ad opera dell'Appendix B della direttiva cfr. *supra*, p. 182 n. 594.

valori utilizzati come criterio di giudizio[740] appaia inesistente o moderato – non sono previste garanzie e risultano leciti, quindi, processi decisionali senza intervento umano diretto. Quando il rischio appaia alto o molto alto (Level III-IV della tassonomia definite dall'Appendix B), invece, l'Appendix C prevede requisiti stringenti, imponendo che «decisions cannot be made without having specific human intervention points during the decision-making process; and the final decision must be made by a human».

Esaurito l'inquadramento dei principali esempi di positivizzazione del diritto al controllo umano sulla tecnologia, preme evidenziare che le regolazioni europea e canadese condividono un limite evidente, già visto in materia di diritto alla spiegazione: limitano il loro campo d'applicazione all'ambito della decisione automatizzata riguardante la persona fisica, peraltro solo se amministrativa, nel caso canadese, e solo quando si basi su un trattamento di dati personali, nel caso europeo[741]. Quasi non si contano le applicazioni di tecnologie avanzate, in primo luogo l'intelligenza artificiale, che non rientrano in questa categoria, ma che paiono idonee a rappresentare un rischio per una pluralità di interessi individuali, di fronte al quale il riconoscimento del diritto allo *human oversight* potrebbe rappresentare una tutela efficace. Basti pensare agli esempi, già citati, della robotica industriale[742], degli strumenti tecnologici impiegati nella gestione di situazioni ad alto rischio[743], dei veicoli totalmente o parzialmente autonomi[744], o, per quanto riguarda i confini della normativa europea, a ogni elaborazione automatizzata di dati non rientranti nella definizione di dati personali (ad esempio perché anonimizzati)[745].Dunque, per quanto il riconoscimento del diritto a uno *human in the loop* appaia, specialmente nel caso del GDPR, molto

[740] Si tratta, come già riportato, di «the rights of individuals or communities, the health or well-being of individuals or communities, the economic interests of individuals, entities, or communities, the ongoing sustainability of an ecosystem».

[741] Cfr. *supra*, p. 209 ss.

[742] Cfr. ancora A.F.T. WINFIELD ET AL., *Robot Accident Investigation: A Case Study in Responsible Robotics cit.*

[743] A. OSUNWUSI, *Aviation Automation and CNS/ATM-related Human-Technology Interface: ATSEP Competency Considerations cit.*; Y. CHENG ET AL., *Reliability Prediction and Safety Evaluation of ATC Automation System cit.*

[744] Il mantenimento di un adeguato livello di controllo umano sui veicoli caratterizzati da un grado avanzato di autonomia è una delle questioni più dibattute tra gli specialisti del settore, a causa della necessità di contrastare fattori come la tendenza alla distrazione e la stanchezza del soggetto chiamato non più a guidare, ma semplicemente a sorvegliare il sistema e intervenire in caso di emergenza. Deve evidenziarsi che non mancano, nel dibattito, le voci che vedono nello sviluppo di veicoli completamente autosufficienti, in cui l'intervento umano non sia previsto, l'obiettivo da raggiungere in materia di sicurezza. Cfr., da vari punti di vista, M. L. CUNNINGHAM, M. A. REGAN, *Driver distraction and inattention in the realm of automated driving*, in *IET Intelligent Transport Systems*, 12, 6, 2018; N. MERAT ET AL., *The "Out-of-the-Loop" concept in automated driving: proposed definition, measures and implications*, in *Cognition, Technology & Work*, 21, 1, 2019; T. L. LOUW, N. MERAT, A. H. JAMSON, *Engaging with highly automated driving: To be or not to be in the loop?*, in *8th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Leeds-Salt Lake City, 2015, https://eprints.whiterose.ac.uk/84892/ (10 agosto 2022); T. LOUW, R. MADIGAN, O. CARSTEN, N. MERAT, Were they in the loop during automated driving? Links between visual attention and crash potential, in *Injury Prevention*, 23, 4, 2017; J. WU ET AL., *Human-in-the-Loop Deep Reinforcement Learning with Application to Autonomous Driving*, arXiv, 2021, https://arxiv.org/abs/2104.07246 (10 agosto 2022).

[745] Sulle questioni sollevate dall'impiego di *database* anonimizzati per lo sviluppo di algoritmi predittivi e sul possibile vuoto di tutela a riguardo, generato dal campo di applicazione dell'attuale normativa europea sui dati personali, si rimanda ancora alle considerazioni svolte *supra*, p. 211 ss. e n. 699.

più netto di quello del diritto alla spiegazione (banalmente, l'espressione «diritto di ottenere l'intervento umano» è presente nell'articolato del Regolamento) l'ambito applicativo di tale nuova posizione giuridica appare del tutto inadeguato. Ciò è una diretta conseguenza delle finalità per cui sono stati concepiti i testi normativi in cui il riconoscimento del diritto è inserito nei due ordinamenti presi in esame, quello europeo e quello canadese. Né il GDPR né la direttiva canadese, infatti, sono stati concepiti per regolare, ad ampio spettro, l'intelligenza artificiale e l'interazione con essa dell'essere umano, e le disposizioni in materia in essi contenute si espongono, inevitabilmente, a pesanti vuoti di tutela.

4.4. Il diritto al controllo umano sul sistema come diritto fondamentale alla luce del quadro giuridico attuale e della Proposta di Regolamento dell'Unione Europea

Al netto del loro ambito d'applicazione eccessivamente ristretto, è doveroso evidenziare che le modalità in cui, sia nell'Unione Europea che in Canada, è stato positivizzato il requisito dello *human oversight* paiono, invece, apprezzabili. Entrambe le formulazioni, infatti, enunciano, come principio generale, la necessità di garantire un intervento umano, per poi limitarsi a tratteggiarne nelle linee essenziali le modalità di attuazione concreta. A una statuizione non priva di una certa solennità (il GDPR, in particolar modo, si esprime esplicitamentein termini di *diritto* all'intervento umano) si accompagna una snella regolazione di dettaglio, in grado di garantire che la posizione giuridica in questione non venga compressa oltre il suo nucleo essenziale e conservi, al contempo, la necessaria flessibilità applicativa. Un'enunciazione di principio che sembra particolarmente adatta al riconoscimento al requisito dello *human in the loop* dello *status* di diritto fondamentale.

L'elevazione del controllo umano al rango di diritto, del resto, appare agevole sesi considera il suo valore metagiuridico. La sottoposizione di ogni applicazione tecnologica al controllo dell'essere umano, infatti, attiene alla visione che l'uomo ha del proprio rapporto con il mondo esterno, e in particolare la componente artificiale di quest'ultimo. Come anticipato nella trattazione del diritto alla spiegazione, non può negarsi che tale rapporto sia stato concepito, fin dalle epoche più remote, in modo chiaramente antropocentrico[746]. La garanzia che tale concezione basilare sia preservata, con la predisposizione di apposite strategie per il controllo, la sorveglianza, e l'eventuale intervento dell'essere umanonel funzionamento delle tecnologie avanzate, allora, non può che chiamare in causa le categorie giuridiche più basilari. È anche per questo consolidato retroterrafilosofico e culturale, forse, che il Legislatore europeo non ha esitato a utilizzare, nelle disposizioni vincolanti del GDPR, il termine "diritto" in riferimento al controllo umano: l'unico caso in cuiuna delle tre nuove posizioni giuridiche qui esaminate ha già trovato tal genere di riconoscimento.

[746] Cfr. *supra*, p. 189 ss.

Svolte queste considerazioni, è bene, in chiusura di questo capitolo, soffermarsi sulle norme riguardanti i requisiti tecnici volti a garantire la presenza di uno *human in the loop* previste dalla già più volte citata Proposta di Regolamento in materia di intelligenza artificiale, presentata dalla Commissione nell'aprile 2021 e ora allo studio delle istituzioni europee. Tale ipotesi di regolazione, infatti, da un lato pare in grado di porre rimedio alla vista settorialità della disciplina oggi in vigore nell'Unione Europea, limitata all'ambito di applicazione della decisione automatica, dall'altro appare particolarmente onerosa e dettagliata, tanto da aver portato alcuni addettti ai lavori a esprimere dubbi sulla sua concreta realizzabilità[747].

A venire in gioco è, in particolare, l'art. 14 della Proposta, eloquentemente rubricato, nella versione in lingua inglese, *human oversight* (reso, nella traduzione italiana, con *sorveglianza umana*) e da applicarsi, in caso di approvazione, a tutte le tecnologie considerate ad alto rischio (categoria, com'è noto, estremamente ampia)[748]. La norma, al paragrafo 2, identifica le funzioni della

[747] Significative, da questo punto di vista, sembrano in particolare le affermazioni di Google nel *position paper* trasmesso alla Commissione Europea durante la consultazione pubblica sulla Proposta di Regolamento condotta nella seconda metà del 2021, consultabile all'indirizzo https://www.ai.google/responsibilities/public-policy-perspectives/ (8 agosto 2022): «Article 14(4)(a) requires that individuals that exercise human oversight of AI systems "fully understand the capacities and limitations of the high-risk AI system." For many AI systems, whether highly complex models with millions or billions of parameters or relatively simple hand-coded models, "fully understanding" the system is effectively impossible. Rather individuals should be required to "adequately understand" the system to exercise effective oversight. Furthermore, while this requirement will necessitate appropriate documentation from the developer of the system, as noted above, the deployer, rather than the developer of the system, will need to ensure that the individual performing oversight has the appropriate understanding», p. 9. Il documento, inoltre, contiene un nutrito elenco dall'eloquente titolo: «Compliance obligations that will be difficult or impossible for providers of general-purpose AI systems to meet under most circumstances» (p. 5). Si vedano anche le osservazioni, presentate nello stesso contesto, del *think-tank* Center for Data Innovation, https://www2.datainnovation.org/2021-feedback-aia.pdf (8 agosto 2022). Altre prese di posizione dello stesso tenore possono consultarsi tra i risultati della consultazione, resi pubblici dalla Commissione: https://bit.ly/3qsHSRV (8 agosto 2022). Si rimanda, inoltre, ai già citati DIGITALEUROPE, *Report - DigitaleuropÈs Initial Findings on the Proposed AI Act cit.*;P. GLAUNER, *An Assessment of the AI Regulation Proposed by the European Commission cit.*

[748] L'art. 14 della Proposta recita: «1. I sistemi di IA ad alto rischio sono progettati e sviluppati, anche con strumenti di interfaccia uomo-macchina adeguati, in modo tale da poter essere efficacemente supervisionati da persone fisiche durante il periodo in cui il sistema di IA è in uso. 2. La sorveglianza umana mira a prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali che possono emergere quando un sistema di IA ad alto rischio è utilizzato conformemente alla sua finalità prevista o in condizioni di uso improprio ragionevolmente prevedibile, in particolare quando tali rischi persistono nonostante l'applicazione di altri requisiti di cui al presente capo. 3. La sorveglianza umana è garantita mediante almeno una delle seguenti misure: a) misure individuate e integrate nel sistema di IA ad alto rischio dal fornitore prima della sua immissione sul mercato o messa in servizio, ove tecnicamente possibile; b) misure individuate dal fornitore prima dell'immissione sul mercato o della messa in servizio del sistema di IA ad alto rischio, adatte ad essere attuate dall'utente. 4. Le misure di cui al paragrafo 3 consentono le seguenti azioni, a seconda delle circostanze, alle persone alle quali è affidata la sorveglianza umana: a) comprendere appieno le capacità e i limiti del sistema di IA ad alto rischio ed essere in grado di monitorarne debitamente il funzionamento, in modo che i segnali di anomalie, disfunzioni e prestazioni inattese possano essere individuati e affrontati quanto prima; b) restare consapevole della possibile tendenza a fare automaticamente affidamento o a fare eccessivo affidamento sull'output prodotto da un sistema di IA ad alto rischio ("distorsione dell'automazione"), in particolare per i sistemi di IA ad alto rischio utilizzati per fornire informazioni o raccomandazioni per le decisioni che devono essere prese da persone fisiche; c) essere in grado di interpretare correttamente l'output del sistema di IA ad alto rischio, tenendo conto in particolare delle caratteristiche del sistema e degli strumenti e dei metodi di interpretazione disponibili; d) essere in grado di decidere, in qualsiasi situazione particolare, di non usare il sistema di IA ad alto rischio o altrimenti di ignorare, annullare o ribaltare l'output del sistema di IA ad alto rischio; e) essere in grado di intervenire sul funzionamento del sistema di IA ad alto rischio o di interrompere il sistema mediante un pulsante di "arresto" o una procedura analoga. 5. Per i sistemi di IA ad alto rischio di cui all'allegato III, punto 1, lettera a), le misure di cui al

sorveglianza umana, esplicitando il suo stretto collegamento con un'ampia gamma di diritti fondamentali: «la sorveglianza umana mira a prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali che possono emergere quando un sistema di IA ad alto rischio è utilizzato conformemente alla sua finalità prevista o in condizioni di uso improprio ragionevolmente prevedibile». I paragrafi successivi, e in particolare il paragrafo 4, dettano, invece, stringenti adempimenti tecnici, da applicare perché possa dirsi sussistente un appropriato livello di *human oversight*. Apprezzabilmente, è previsto l'obbligo di mettere in atto misure cheavvisino l'utentedel rischio di fareeccessivo affidamento sulla tecnologia; la Proposta, inoltre, impone che nonvengamai meno la possibilità di fermare il funzionamento di un sistema e sovvertirne i risultati (prerogative che, come già detto, potrebbero definirsi il nucleo essenziale del diritto al controllo umano). Altre parti della norma, però, impongono requisiti particolarmente severi, che paiono strettamente legati al grado di trasparenza del sistema. Il menzionato art. 14 par. 4, ad esempio, prevede, tra le altre cose, che l'operatore umano sia messo in grado di «interpretare correttamente l'output [...], tenendo conto, in particolare, delle caratteristiche del sistema e degli strumenti e dei metodi di interpretazione disponibili» o di «intervenire sul funzionamento del sistema». Nonostante queste disposizioni invitino a tenere conto dello stato dell'arte tecnologico – ad esempio quando fanno riferimento ai «metodi di interpretazione disponibili» - esse sembrano richiedere, se rigorosamente interpretate, un livello di controllo in capo all'essere umano che talvolta potrebbe risultare difficilmente applicabile a tecnologie scarsamente interpretabili, come alcune applicazioni delle reti neurali profonde. Anche l'eventuale applicazione di tecniche di *explainable artificial intelligence* particolarmente avanzate potrebbe non essere sufficiente a garantire all'utente la possiblità di interferire attivamente con ogni ciclo di funzionamento del sistema: tale ipotesi, infatti, pare realizzabile solamente qualora gli stati interni del sistema siano comprensibili, e non con l'elaborazione di una spiegazione *ex post* del risultato, variamente ottenuta[749]. Proprio questa constatazione ha portato, come detto, diversi specialisti a sollevare perplessità sulla concreta possibilità di rispettare, con le attuali tecnologie, la norma nella sua interezza[750].

---

paragrafo 3 sono tali da garantire che, inoltre, l'utente non compia azioni o adotti decisioni sulla base dell'identificazione risultante dal sistema, a meno che essa non sia stata verificata e confermata da almeno due persone fisiche».

[749] Sulla distinzione tra tecniche di *explainable artificial intelligence* che puntano all'interpretazione degli stati interni del sistema e tecniche che puntano a una spiegazione degli *output*, cfr. ampiamente *supra*, p. 193ss. In letteratura si rinvia di nuovo a G. MONTAVON, W. SAMEK, K.R. MÜLLER, *Methods for interpreting and under standing deep neural networks cit.*; G. VILONE, L. LONGO, *Explainable artificial intelligence: a systematic review cit.*; Z. C. LIPTON, *The mythos of model interpretability cit.*; T. SPEITH, *A review of taxonomies of expleinable artificial intelligence (XAI) methods cit.*; C. MESKE, E. BUNDE, J. SCHNEIDER, M. GERSCH, *Explainable artificial intelligence: objectives, stakeholders, and future research opportunities cit.*

[750] Si rimanda nuovamente alle osservazioni alla Proposta di Regolamento prodotte da Google, https://www.ai.google/responsibilities/public-policy-perspectives/ (8 agosto 2022) e dal Center for Data Innovation, https://www2.datainnovation.org/2021-feedback-aia.pdf (8 agosto 2022) nel corso nella consultazione pubblica avviata

Dunque, l'intensità dello*human oversight* richiesto sembrerebbe accostabile, utilizzando la tassonomia indicativa teorizzata al capitolo precedente, al livello che le *Linee Guida etiche* del Gruppo di Esperti di alto livello sull'IA della Commissione definiscono *human in the loop*. Ciòpotrebbe portare all'inutilizzabilità, in quanto non conformi al Regolamento, di molte tecnologie – in primo luogo quelle basate sull'apprendimento automatico – che spesso si comportano come *black-box*, senza che, in alcuni casi, esistano alternative interpretabili in grado di svolgere con pari efficacia le stesse funzioni. Sarebbe preclusa ogni valutazione sulla possibilità e legittimità del loro utilizzo in funzione dei valori concretamente coinvolti anche qualora, per ipotesi, essi rientrino tra i più basilari. L'ovvio esempio potrebbe essere rappresentato, ancora una volta, da eventualistrumenti di intelligenza artificiale potenzialmente *life-saving*, come alcune applicazioni delle reti neurali in ambito diagnostico e terapeutico[751]. L'art. 14 della Proposta di Regolamento, così, impedirebbe al diritto allo*human oversight* di prestarsi al bilanciamentocon altri diritti e interessi di pari rango, comprimendosi ove necessario. In ogni caso, è già stata evidenziata la presenza, pur non preponderante, di alcuni elementi di apertura a un'interpretazione flessibile nell'attuale formulazione dell'art. 14 della Proposta. In caso di approvazione, sarà,allora, la lettura che nedaranno giudici, autorità di settore e operatori del mercato a determinare se il regime ivi previsto risulti eccessivamente rigido e difficilmente applicabile o consenta, invece, approcci più flessibili e maggiormente rispondenti alla categoria dei diritti fondamentali, cui, secondo l'orientamento di questo lavoro, il controllo umano sulla tecnologia appartiene.

in proposito dalla Commissione, e ai citati DIGITALEUROPE, *Report – Digital Europe's Initial Findings on the Proposed AI Act cit.*; P. GLAUNER, *An Assessment of the AI Regulation Proposed by the European Commission cit.*

[751]Possono menzionarsi, ad esempio, alcuni sistemi di *image recognition* particolarmente efficaci per il riconoscimento dei tumori della pelle e del pancreas, cfr.ESTEVA, B. KUPREL, R. A. NOVOA, J. KO, S. M. SWETTER, H. M. BLAU, *Dermatologist-level classification of skin cancer with deep neural networks*, in *Nature*, 542, 7639, 2017, p. 115-118; S. MATHOTARACHI, M. ZHU, C. XU, J. YU, Y. WU, C. LI, M. ZHANG*, Differentiation of Pancreatic Cancer and Chronic Pancreatitis Using Computer-Aided Diagnosis of Endoscopic Ultrasound (EUS) Images: A Diagnostic Test*, in *PLoS ONE*, 8, 5, 2013; o alcune applicazioni dell'apprendimento automatico per l'identificazione precoce di determinati fattori di rischio o malattie neurodegenerative, cfr. S. F. WENG, J. REPS, J. KAI, J. M. GARIBALDI, N. QURESHI, *Can machine-learning improve cardiovascular risk prediction using routine clinical data?*, in *PLOS ONE*, 12, 4, 2017; T. A. PASCOAL, M. SHIN, A. L. BENEDET, M. KANG, T. BEAUDRY, *Identifying incipient dementia individuals using machine learning and amyloid imaging*, in *Neurobiology of Aging*, 59, 2017, p. 80-90. Riguardo all'applicazione dell'intelligenza artificiale in ambito medico cfr. ampiamente *infra*, p. 277 ss.

I nuovi diritti messi alla prova. L'intelligenza artificiale nell'attività amministrativa, giudiziaria e medica

1. Intelligenza artificiale, Pubblica Amministrazione e diritti fondamentali

1.1. Cenni sui principali utilizzi dell'intelligenza artificiale da parte delle Pubbliche Amministrazioni dei paesi democratici

La varietà di possibili applicazioni dell'intelligenza artificiale non ha, ovviamente, attratto l'interesse dei soli operatori privati. Al contrario, i poteri pubblici di ogni paese avanzato, almeno a partire dagli anni 2000, hanno incorporato nelle loro attività un numero crescente di tecnologie basate sull'intelligenza artificiale. La trasformazione ha riguardato pressochè ogni ambito dell'azione statale: è il caso, ad esempio, dell'apparato militare – l'impiego dell'intelligenza artificiale per la costruzione di armamenti intelligenti, peraltro, suscita notevoli preoccupazioni[752] – del sistema giudiziario [753] , come si vedrà nei paragrafi successivi, o della pubblica amministrazione[754], sullaquale ora si concentrerà l'analisi. Quest'ultima parte del lavoro, infatti, è

[752] Suscitano preoccupazione, in particolare, i c.d. *autonomous weapons system*, strumenti bellici spesso manovrabili a distanza o totalmente automatizzati, in grado di minimizzare il rischio di perdite per chi li impieghi, a causa del pericolo di una disumanizzazione ulteriore del contesto bellico che potrebbe conseguirne. Cfr. D. AMOROSO, *Jus in bello and jus ad bellum arguments against autonomy in weapons systems: a re-appraisal*, in *Questions on International Law,* Zoom in 43, 2017, p. 5-31; M. SASSÓLI, *Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified*, in *International law studies*, 90, 2014, p. 308-340; P. ASARO, *On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making*, in *International Review of the Red Cross*, 2012, 94, p. 687-709; F. CHESINI, *"Terminator scenario"? Intelligenza artificiale nel conflitto armato: "lethalautonomousweapons systems" e le risposte del diritto internazionale umanitario*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2020, p. 441-471.

[753] Tra i moltissimi contributi sulle applicazioni dell'intelligenza artificiale in ambito giudiziario si indicano fin d'ora C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi cit.*, F. DONATI, *Intelligenza artificiale e giustizia*, in *Rivista AIC*, 1, 2020; A. GARAPON, J. LASSÈGUE, *Justice digitale,* Parigi, 2018; M. LUCIANI, *La decisione giudiziaria robotica*, in *Rivista AIC*, 3, 2018, p. 872-893;A. SANTOSUOSSO, *Intelligenza Artificiale e Diritto : Perché Le Tecnologie Di IA Sono Una Grande Opportunità per Il Diritto* , Milano, 2020; A. GARAPON, J. LASSEGUE, *Justice digitale: révolution graphique et rupture anthropologique*, Parigi,2018.M. FASAN, *L'intelligenza artificiale nella dimensione giudiziaria. Primi profili giuridici e spunti dall'esperienza francese per una disciplina dell'AI nel settore della giustizia* in*Gruppo di Pisa*, Quaderno monografico n. 3, 2021, 325-339; S. PENASA,*Intelligenza artificiale e giustizia: il delicato equilibrio tra affidabilità tecnologica e sostenibilità costituzionale in prospettiva comparata*, in *DPCE Online*, 1, 2022, p. 297-310;M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessmenttools tra Stati Uniti ed Europa*, in*Diritto Penale Contemporaneo*,2019,https://bit.ly/3UsvzmD(21 giugno 2022);D. POLIDORO, *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dall'intelligenza artificiale*, in *Archivio Penale*, 3,2020; U. PAGALLO – S. QUATTROCOLO, *The impact of AI inCriminal Law, and itsTwofoldProcedures*, in W. BARFIELD- U. PAGALLO (a cura di) ,*Research Handbook on the Law of Artificial Intelligence*, Cheltenam, 2018, 388 ss.; S. QUATTROCOLO, *Equo processo penale e sfide della società algoritmica*, in *Rivista di BioDiritto – BioLaw Journal*, 1, 2019, p. 135 ss.;S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs. rischi e paure della giustizia digitale "predittiva"*, in *Cassazione penale*, 59, 4, 2019, p. 1748 ss.; D. D. LUXTON, *Should Watson Be Consulted for a Second Opinion?*, in*AMA Journal of Ethics*, 2, 2019, p. 131-138; T. SOURDIN, *Judge v. Robot? Artificial Intelligence and Judicial decision-making*, in *UNSW Law Journal*, 4, 2018, p. 1114-1133; A. ZAVRŠNIK, *Criminal justice, artificial intelligence systems, and human rights*, in *ERA Forum*, 20, 4, 2020;W. S. ISAAC, HOPE, HYPE E FEAR*, The Promise and Potential Pitfalls of Artificial Intelligence in Criminal Justice,* in *Ohio State Journal of Criminal Law,* 15, 2017.

[754]La letteratura riguardante le applicazioni dell'intelligenza artificiale nella pubblica amministrazione è vastissima. *Ex multis* si rimanda innanzitutto a B. MARCHETTI, *La garanzia dello human in the loop alla prova della decisione*

dedicata all'approfondimento delle questioni poste, in materia di diritti fondamentali, dall'applicazione dell'intelligenza artificiale in tre ambiti di specifici: la pubblica amministrazione, la giustizia e l'attività medico-sanitaria. La scelta è caduta su questi tre domini, tra i molti possibili, innanzitutto in ragione della rilevanza dei diritti coinvolti. I settori della pubblica amministrazione e della giustizia, infatti, sono i campi in cui, più che in ogni altro, si esprime la relazione tra individuo e potere, come già detto al cuore di ogni diritto fondamentale. Il rilievo delle posizioni giuridiche chiamate in causa dall'attività medica, invece, è autoevidente: a venire in gioco sono, in tal caso, i beni della vita, della salute e dell'integrità fisica, la cui garanzia è il presupposto per l'esercizio e il godimento di tutti gli altri diritti e libertà. Per quanto riguarda l'ambito medico, inoltre, a renderlo di particolare interesse ai fini di questo studio è anche, come si dirà, la particolare apertura all'innovazione tecnologica che lo caratterizza, che rende le tecnologie intelligenti – e le questioni da esse sollevate – particolarmente diffuse nel settore.

Del resto, la stessa letteratura scientifica ha dedicato particolare attenzione alle applicazioni dell'intelligenza artificiale in questi ambiti. Ciò è vero, innnanzitutto, per quanto riguarda il coinvolgimento di sistemi avanzati nelle attività della pubblica amministrazione. Infatti, studiosi di diritto pubblico, costituzionale e amministrativo ne hanno commentato estensivamente l'utilizzo, vedendovi, allo stesso tempo, notevoli opportunità di incrementare l'efficienza degli apparati statali e rischi inediti per le tutele normalmente garantite all'individuo nei rapporti con le autorità degli ordinamenti democratici. La riflessione scientifica si è concentrata soprattutto sull'impiego di intelligenza artificiale nel procedimento amministrativo[755]. Ciò è la conseguenza dell'attenzione

---

amministrativa algoritmica, in *BioLaw Journal - Rivista di BioDiritto*, 2, 2021; RAFFIOTTA E., *L'erompere dell'intelligenza artificiale per lo sviluppo della pubblica amministrazione e dei servizi al cittadino*, in G.C. FERONI, C. FONTANA, E.C. RAFIOTTA, *AI ANTHOLOGY - Profili giuridici, economici e sociali dell'intelligenza artificiale*, Bologna, 2022; E. CHITI; B. MARCHETTI; N. RANGONE, *L'impiego di sistemi di intelligenza artificiale nelle pubbliche amministrazioni italiane: prove generali*, in*BioLaw Journal - Rivista di BioDiritto*, 2, 2022; A. PAJNO E AL.,*AI: profili giuridici. Intelligenza artificiale: criticità emergenti e nuove sfide per i giuristi*, in *Biolaw Journal*, 3, 2019, 205; 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, 2, 2019; G. AVANZINI, *Decisioni amministrative e algoritmi informativi. Predeterminazione analisi predittiva e nuove forme di intellegibilità*, Napoli, 2019;M. BASSINI, L. LIGUORI, O. POLLICINO, *Sistemi di intelligenza artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, in F. PIZZETTI (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018;L. CASINI, *Lo Stato nell'era di Google. Frontiere e sfide globali*, Milano, 2020;A. SIMONCINI, *Profili costituzionali della amministrazione algoritmica*, in *Rivista trimestrale di diritto pubblico*, 4, 2019; C. COGLIANESE, D. LEHR, *Transparency and AlgorithmicGovernance*, in *Administrative Law Review*, 71, 7, 2019; D. VALLE-CRUZ; E. ALEJANDRO RUVALCABA-GOMEZ; W. G. DE SOUSA; E. R. P. DE MELO; P. H. D. S. BERMEJO; R. A. S. FARIAS; A. O. GOMES, *How and where is artificial intelligence in the public sector going? A literature review and research agenda*, in *Government Information Quarterly*, vol. 36, 4, 2019.

[755]Pressochè ogni contributo citato alla nota precedente tratta approfonditamente il tema. Cfr. inoltre, nella letteratura giuridica italiana, N. MUCIACCIA, *Algoritmi e procedimento decisionale: alcuni recenti arresti della giustizia amministrativa*, in *Federalismi.it*, 15 aprile 2020, https://bit.ly/3DGB0bH (22 agosto 2022); A. SIMONCINI, *Amministrazione digitale algoritmica. Il quadro costituzionale*, in D. U GALETTA, R. CAVALLO PERIN (A CURA DI), *Il diritto dell'Amministrazione Pubblica digitale,* Torino, 2020, p. 1-41; F.LAVIOLA, *Algoritmico, troppo algoritmico: decisioni amministrative automatizzate, protezione dei dati personali e tutela delle libertà dei cittadini alla luce della più recente giurisprudenza amministrativa*, in *Biolaw Journal*, 3, 2020, p. 389 ss.; A.NICOTRA, V. VARONE,

maggiore che gli studiosi di diritto hanno, come già evidenziato, dedicato all'impiego di strumenti intelligenti nei procedimenti decisionali, specialmente qualora i loro effetti riguardino esseri umani. Le ragioni di tale interesse paiono più che valide, e sono state già analizzate nel corso del lavoro: l'automazione, totale o parziale, di scelte e valutazioni in precedenza esclusiva dell'essere umano sembra erodere la stessa concezione antropocentrica del rapporto tra uomo e macchina; l'esposizione di uno o più individui alle conseguenze di scelte altrui è sempre accompagnata da tutele giuridiche, la cui efficacia pare messa in dubbio dall'entrata in gioco dell'elemento artificiale. Tale ultima circostanza è veritiera, a maggior ragione, per la decisione amministrativa, che nei sistemi democratici, pur con importanti differenze, è sempre accompagnata da corpose garanzie procedimentali e sostanziali [756]. Inoltre, proprio l'azione di soggetti che si ritenevano lesi dall'utilizzo di algoritmi in valutazioni dei poteri pubblici ha portato, come si approfondirà, l'impiego di intelligenza artificiale nelle attività della pubblica amministrazione di fronte alle corti di alcuni paesi. La particolare attenzione dimostrata dalla letteratura giuridica per il coinvolgimento di sistemi intelligenti nella decisione amministrativa, dunque, non pare da biasimare. Ciò nonostante, sarebbe un notevole errore pensare che l'impiego crescente di tecnologie avanzate da parte delle pubbliche applicazioni di vari paesi che ha caratterizzato gli ultimi anni riguardi solamente tale ambito di applicazione. Nel tentativo di offrire una panoramica sintetica delle applicazioni dell'intelligenza artificiale nella pubblica amministrazione ulteriori rispetto al diretto coinvolgimento in procedimenti decisionali o valutativi, possono menzionarsi:

- ***Chatbot* per la gestione delle relazioni col pubblico su larga scala**, basati su tecniche di elaborazione del linguaggio naturale. È ormai estremamente comune l'utilizzo di *chatbot*per facilitare il reperimento di informazioni da parte degli utenti nei siti web di settori della pubblica amministrazione noti per essere diffusamente percepiti come burocratici, scarsamente accessibili e disfunzionali. La diffusione di questi strumenti è avvenuta assieme alla digitalizzazione, totale o parziale, di un numero crescente di attività, come il completamento di adempimenti fiscali o previdenziali e l'iscrizione a concorsi e selezioni pubbliche. Le pagine web degli enti pubblici con cui il cittadino si interfaccia per tali pratiche spesso offrono all'utente la possibilità di interagire col *chatbot*già nella prima

L'algoritmo, intelligente ma non troppo, in *Rivista AIC*, 4, 2019, p. 86 ss.; L. M. AZZENA, *L'algoritmo nella formazione della decisione amministrativa: l'esperienza italiana cit.*Per alcune prospettive su ordinamenti stranieri v. invece, *ex multis*, L. ANDREWS, *Public administration, public leadership and the construction of public value in the age of the algorithm and 'big data'*, in *Public Administration*, 97, 2, 2019;C. COGLIANESE, D. LEHR, *Regulating by robot: Administrative Decision Making in the Machine Learning Era*, in *Georgetown Law Journal*, 2017, 105, p. 1147 ss.; H. G. VAN DERVOORT; A. J. KLIEVINK; M. ARNABOLDI; A. J. MEIJER, *Rationality and politics of algorithms. Will the promise of big data survive the dynamics of public decision making*, in *Government Information Quarterly*, 36, 1, 2019.

[756]Cfr., tra i molti possibili, V. DE FALCO, *Azione amministrativa e procedimenti nel diritto comparato*, Padova, 2018; S. CASSESE, *La disciplina legislativa del procedimento amministrativo. Una analisi comparata*, in *Il Foro Italiano*, 116, 1, 1993, p. 27-34.

schermata. Esistono evidenze scientifiche dell'efficacia di questi strumenti, che semplificano e velocizzano l'interazione degli utenti con la pubblica amministrazione per la maggior parte delle finalità per le quali essi ne utilizzano più di frequente i portali online[757]. Si tratta di sistemi in uso in vari paesi, Italia compresa: il sito dell'INPS, ad esempio, offre ai suoi visitatori il supporto dell'assistente virtuale *Arianna*[758]; l'ente previdenziale francese URSSAF fornisce un servizio analogo[759]; l'*Australian Taxation Office* propone ai suoi utenti l'assistenza di un *chatbot* chiamato Alex[760]. Su queste tecnologie, inoltre, sembra puntare particolarmente il Regno Unito, che ha annunciato l'intenzione di sviluppare, nei prossimi anni, *chatbot* particolarmente sofisticati per facilitare l'interazione tra cittadini ed enti pubblici e migliorare i loro servizi, in particolare nel sistema sanitario nazionale[761].

- **Applicazioni dell'intelligenza artificiale volte a rendere più efficiente l'erogazione di servizi pubblici e l'accesso a beni comuni**, basate su tecnologie di IoT. Alcuni progetti d'avanguardia sono stati avviati, ad esempio, nelle città statunitensi di Baltimora, New York, Washington e San Francisco, in cui l'utilizzo di sensori per il monitoraggio dei flussi idrici permette di identificare e correggere, con inedita precisione e rapidità, perdite e congestioni nei rispettivi acquedotti, portando a una gestione più efficiente delle risorse[762]. Il tema richiama, in una prospettiva più generale, il fenomeno delle *smart cities*, i cui sviluppi futuri sono appena agli inizi: città in cui tecnologie avanzate, e in primo luogo la combinazione di *Internet of Things* e analisi dei dati, sono impiegate per lo sviluppo di forme di regolazione particolarmente efficiente della vita collettiva, basate in tutto o in parte sull'automazione[763]. Si tratta di progetti che, spesso, generano modalità di collaborazione

[757] Cfr. ad esempio K. NIRALA, N. K. SINGH EV.S. PURANI, *A survey on providing customer and public administration-based services using AI: chatbot*, in *Multimed Tools Appl* 81, 2022, p. 22215–22246; T. MAKASI, A. NILI, T. ALIREZA, K. DESOUZA, M. TATE, *Chatbot-mediated public service delivery: a public service value-based framework*, in *First Monday*, 25, 12, https://eprints.qut.edu.au/204999/ (22 agosto 2022).

[758] Si rinvia alla pagina web dell'ente: https://bit.ly/3fbaBs1 (22 agosto 2022).

[759] Il *chatbot* appare nell'homepage del sito dell'ente: https://www.urssaf.fr/portail/home.html (22 agosto 2022), e accoglie gli utenti col messaggio: «Bonjour, je suis là pour vous aider et vous accompagner».

[760] Anche in questo caso, il servizio è disponibile nell'homepage del sito istituzionale: https://www.ato.gov.au/ (22 agosto 2022). Con il primo messaggio orienta gli utenti verso i quesiti più richiesti quel giorno:«Hi, I'm Alex, the ATO's virtual assistant. So I can provide you the best tax and super support, here are the most common topics I've been asked about today [segue l'elenco delle ricerche più ricorrenti]».

[761] Cfr. UK GOVERNMENT - DEPARTMENT OF HEALTH AND SOCIAL CARE, *£36 million boost for AI technologies to revolutionise NHS care*, 16 giugno 2021.

[762] Cfr. M. J. AHN; Y.C. CHEN, *Artificial intelligence in government: potentials, challenges, and the future*, in *The 21st annual international conference on digital government research*, Seoul, 15-19 giugno 2020, https://bit.ly/3LAogoS(18 agosto 2022), p. 243 ss.

[763] Cfr. *ex multis* G. F. FERRARI (A CURA DI), *Le smart cities al tempo dellaresilienza*, Milano, 2021; F. MENEGHETTI, C. ROSSI CHAUVENET, G. FIORONI, *Rapporto 3/2022 – SMART cities e intelligenza artificiale*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2022, p. 253 ss; E. CHITI, B. MARCHETTI, N. RANGONE, *L'impiego di sistemi di intelligenza artificiale nelle pubbliche amministrazioni italiane: prove generali cit.* p. 496 ss.; J.B. AUBY, V. DE GREGORIO, *Le smart cities in Francia. Istituzioni del federalismo*, in *Rivista di studi giuridici e politici*, 4, 2015, p.975-993; H. ARASTEH ET AL., *Iot-based smart cities: A survey*, in 2016 *IEEE 16th International Conference on Environment and Electrical Engineering*

inedite tra attori privati e poteri pubblici. Può farsi l'esempio di iniziative volte a diminuire il consumo energetico degli uffici tramite l'installazione di sensori che regolano, sulla base dell'analisi dei dati, l'accensione e lo spegnimento di luci e dispositivi elettronici, come nel progetto *Smart building management system*, di recente sperimentato ad Amsterdam[764]; di piattaforme di *data-sharing* a finanziamento pubblico contenenti dati anonimizzati su trasporti, condizioni di vita e abitudini di spesa degli abitanti, al fine di permettere a ricerca e industria di proporre e sviluppare soluzioni *smart*, come nel caso del *Helsinki Region Infoshare* avviato nel 2013 nella capitale della Finlandia[765]; di strategie per il miglioramento del livello di sicurezza stradale, della gestione del traffico e dell'inquinamento causato dalla circolazione di veicoli come il progetto *The Copenhagen Wheel*, che prevede la possibilità, per gli abitanti della città, di installare sensori sulle ruote della propria bicicletta per trasmettare alle autorità comunali dati anonimizzati sulle condizioni del viaggio. Le informazioni raccolte sono utilizzate per perfezionare la rete di piste ciclabili, disincentivare l'uso dell'automobile e migliorare la qualità dell'aria nei punti che risultino più inquinati[766].

- **Sistemi di intelligenza artificiale impiegati con funzioni di monitoraggio e controllo**, spesso basati sull'apprendimento automatico. L'elaborazione di grandi moli di dati con tecnologie intelligenti è impiegata, in particolare, da un numero crescente di Autorità amministrative indipendenti di vari paesi, per una pluralità di finalità, come lo sviluppo di modelli predittivi di eventi avversi, l'implementazione di politiche antifrode più efficaci, o il miglioramento dell'attività di controllo di cui siano incaricate. Questa trasformazione riguarda anche le attività di alcune Autorità indipendenti italiane, e in particolare dei principali regolatori dei mercati finanziari. La Banca d'Italia, ad esempio, utilizza strumenti di *machine learning* al fine di prevedere la possiblità di default di imprese italiane, identificare operazioni sospette e valutare possibili rischi di riciclaggioo infiltrazione mafiosa[767]. Consob, invece, sta sperimentando un sistema di elaborazione del linguaggio naturale per l'analisi di testi, con cui selezionare i documenti di illustrazione delle caratteristiche dei prodotti finanziari offerti agli investitori al dettaglio che necessitino la

(EEEIC), 2016, doi: 10.1109/EEEIC.2016.7555867 (21 agosto 2022); S. PELLICER, G. SANTA, A. L. BLEDA, R. MAESTRE, A. J. JARA, A. G. SKARMETA, *A global perspective of smart cities: a survey*, in *Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, 2013, p. 439-444.

[764] Cfr. EUROPEAN PARLIAMENT -DIRECTORATE GENERAL FOR INTERNAL POLICIES, *Study -Mapping smart cities in the EU*, gennaio 2014, https://bit.ly/3faciX1 (20 agosto 2022), p. 143 ss.

[765] Si veda, a riguardo, il sito web del progetto: https://hri.fi/en_gb/ (20 agosto 2022).

[766] Cfr. il sito web dell'iniziativa: http://senseable.mit.edu/copenhagenwheel/index.html (20 agosto 2022).

[767] Per maggiori dettagli cfr. F. FEDERICO, J. MARCUCCI, M. BEVILACQUA, D. J. MARCHETTI, *Rapporto 2/2021 – L'impiego dell'IA nell'attività di Banca d'Italia*, in *BioLaw Journal – Rivista di BioDiritto*, 4, 2021, p. 229 ss.

revisione di un'analista umano[768]. Sullo scenario italiano,merita una menzione anche un'iniziativa portata avanti dall'AGCOM negli anni 2019 e 2020, consistente in un sistema intelligente utilizzato per misurare frequenza e modalità di diffusione di discorsi d'odio sui *social media*, i cui risultati venivano diffusi al pubblico per mezzo dell'Osservatorio sulla disinformazione online[769]. Strumenti di intelligenza artificiale simili a quelli citati sono utilizzati da diverse autorità amministrative di altri paesi dell'Unione Europea, a cominciare da quelle incaricate della vigilanza sul sistema bancario e sui mercati finanziari[770]. Al di fuori dell'attività delle Autorità indipendenti, invece, strumenti di intelligenza artificiale con finalità di monitoraggio sono diffusi soprattutto negli Stati Uniti, in cui l'alto tasso di digitalizzazione di alcuni enti locali permette l'impiego, anche su scala cittadina, di questo genere di tecnologie. È il caso, ad esempio, della città di Chicago, in cui l'utilizzo di sistemi di apprendimento automatico per lo sviluppo di modelli predittivi ha permesso di migliorare del 20% circa l'efficienza dei controlli finalizzati alla prevenzione di frodi alimentari, con una tecnologia poi imitata anche dalla città di Boston[771].

- **Sistemi di intelligenza artificiale per ottimizzare l'utilizzo di risorse pubbliche**, in cui l'analisi dei dati è utilizzata al fine di identificare sprechi e inefficienze e prevedere bisogni futuri, al fine di collocare al meglio risorse limitate[772]. Si tratta di applicazioni tecnologiche sempre più diffuse per razionalizzare la gestione di servizi pubblici riconducibili al *welfare state*, in particolare negli Stati Uniti d'America. Come si dirà, alcuni di questi strumenti paiono meritevoli di un'analisi approfondita, per la possibilità che celino finalità non dichiarate di riduzione della spesa, o portino a una diminuzione della tutela di alcuni diritti sociali[773]. Tra gli esempi virtuosi, può menzionarsi un'iniziativa condotta dal *Cincinnati*

[768] Cfr., sul punto, E. CHITI, B. MARCHETTI, N. RANGONE, T. TOGNA, A. LIMOSANI, G. FREGA, P. DERIU, G. RAGUCCI, L. LO SCHIAVO, L. LAZZA, *Rapporto 1/2021 – L'impiego dell'IA nell'attività di CONSOB, AGCOM e ARERA*, in *BioLaw Journal – Rivista di BioDiritto*, 4, 2021, p. 211 ss.

[769] Cfr. in particolare E. CHITI, B. MARCHETTI, N. RANGONE, T. TOGNA, A. LIMOSANI, G. FREGA, P. DERIU, G. RAGUCCI, L. LO SCHIAVO, L. LAZZA, *Rapporto 1/2021 – L'impiego dell'IA cit.*, p. 222 ss.

[770] Sul punto, v. ad es. A. FERNANDEZ, *Artificial Intelligence in Financial Services*, Banco de Espana – Economic Bulletin, 2, 2019.

[771]Cfr. M. J. AHN; Y. C. CHEN, *Artificial Intelligence in Government: Potentials, Challenges, and the Future cit.*, p. 247 ss.; S. THORTON, *Delivering Faster Results with Food Inspection Forecasting*, in *Data-smart city solutions*, 19 maggio 2015, https://bit.ly/3qZluA4 (20 agosto 2022).

[772]È il caso, ad esempio, dei già citati H.H. ARCOLEZI, J.F. COUCHOT, S. CERNA, C. GUYEUX, G. ROYER, B.A. BOUNA ET AL., *Forecasting the number of firefighter interventions per region with local-differential-privacy-based data cit.*; Y. FENG, Q. DUAN, X. CHEN, S.S. YAKKALI, J. WANG, *Space cooling energy usage prediction based on utility data for residential buildings using machine learning methods cit*.

[773] Gli strumenti in esame, infatti, incidono su quelle che G. CALABRESI, P. BOBBITT, *Tragic Choices cit.* definiscono *first-order choices*: scelte allocative che determinano l'assetto delle risorse da cui deriva la necessità delle atomistiche *second-order choices*, ovvero le decisioni su chi, caso per caso, debba accedere a una determinata risorsa scarsa. L'impatto sui diritti, il contenuto etico e la stessa esistenza di tali scelte risultano, spesso, difficilmente identificabili, e il coinvolgimento di sistemi intelligenti, senza le cautele opportune, potrebbe rappresentare un ulteriore elemento di opacità, in ragione, in primo luogo, della diffusa percezione di oggettività che accompagna la tecnologia. Cfr. più approfonditamente *infra*, p. 290 ss.

Fire Department nel 2016, in collaborazione con l'Università di Chicago, che, grazie all'utilizzo di strumenti di intelligenza artificiale nell'analisi dei dati relativi alle chiamate d'emergenza al servizio di pronto soccorso, ha permesso di ridurre del 22% i casi di trasporto tardivo in ambulanza di pazienti critici all'ospedale[774].

1.2 L'impatto dell'amministrazione algoritmica sui diritti fondamentali, vecchi e nuovi

È evidente che l'utilizzo dell'intelligenza artificiale da parte della Pubblica Amministrazione può avere un impatto significativo su una vasta gamma di diritti fondamentali, inclusi quelli identificati come *nuovi diritti* in questo lavoro. Ad esempio, il crescente uso di *chatbot* e assistenti vocali nell'interazione col pubblico pare, in termini generali, da accogliere con favore, perché di dimostrata efficacia nell'agevolare la vita quotidiana dei cittadini. Allo stato dell'arte, inoltre, il settore pubblico sembra rappresentare un esempio virtuoso per ciò che riguarda i problemi sollevati dalla distinguibilità di tali strumenti dall'essere umano. Le applicazioni impiegate dalle pubbliche amministrazioni, infatti, chiariscono quasi sempre la natura artificiale e automatizzata del supporto fornito agli utenti. I*chatbot* menzionati al paragrafo precedente, ad esempio, si presentano all'utente con l'espressione "assistente virtuale". Le questioni principali, da questo punto di vista, sembrano destinate a provenire dal settore privato, in cui, come già visto, *chatbot* e assistenti vocali si stanno diffondendo con grande velocità, e possono esistere interessi specifici, in primo luogo commerciali, a rendere ambigua la loro natura[775]. Tuttavia, non può non evidenziarsi che l'utilizzo di tali sistemi da parte dei poteri pubblici presenta rischi d'altro genere che non paiono trascurabili, il primo dei quali è rappresentato dai possibili esiti discriminatori cui potrebbero condurre[776]. È circostanza nota, infatti, che le competenze digitali variano di molto a seconda di livello culturale, fascia d'età e paese di provenienza. L'Italia, peraltro, in materia appare particolarmente in ritardo, figurando da anni nelle ultime posizioni di ogni classifica che tenti di misurare il livello medio di

[774]Cfr. J. WISEMAN, S. GOLDSMITH, *10 great ways data can make government better*, in *Data-smart city solutions*, 11 maggio 2017, https://bit.ly/2IRpZXX (20 agosto 2022); *Optimizing the Quality and Delivery of City Emergency Medical Services | Data Science for Social Good Fellowship*, in *Data science for social good*, https://bit.ly/3dAtAMu (20 agosto 2022).

[775] Sul punto si rimanda, in primo luogo, ai vantaggi commerciali della mancata *disclosure* evidenziati da X. LUO ET AL., *Machines vs humans: the impact of artificial intelligence chatbot disclosure on customer purchases cit.* e, in senso più ampio, al disagio generato dall'interagire con sistemi antropomorfi realistici e dei quali si conosca la natura artificiale, messi in luce già negli anni '70 da M. MORI, *The uncanny valley cit*. Cfr. più ampiamente *supra*, p. 178 ss.

[776] Oltre alle discriminazioni connesse al c.d. *digital divide*, di seguito analizzate, deve menzionarsi la possibilità che l'implementazione di tali sistemi porti alla discriminazione di alcune minoranze e gruppi svantaggiati, il cui modo di esprimersi non sia compreso da chatbot e assistenti vocali allenati per l'interazione con la lingua "standard". Come visto nella seconda parte, si tratta di vicende già accadute nel diverso contesto della moderazione dei contenuti sui social media, in cui il tasso d'errore di alcuni sistemi automatizzati cresce qualora essi siano utilizzati per l'analisi di alcune varianti della lingua inglese, cfr. S. L. BLODGETT, B. O'CONNOR, *Racial disparity in Natural Language Processing: a case-study of social media African-American English cit.*

alfabetizzazione informatica dei cittadini dei paesi avanzati[777]. La digitalizzazione di un numero crescente di attività della pubblica amministrazione, in assenza di adeguate misure per mitigarne i possibili effetti negativi, potrebbe rendere difficoltoso l'accesso a determinati servizi, anche essenziali, per soggetti in condizione di marginalità, come grandi anziani privi di una solida rete di supporto o persone prive dei mezzi economici per accedere con facilità a strumenti informatici. È agevole obiettare che anche l'interazione di tali categorie con la pubblica amministrazione tradizionale e burocratica, spesso, si svolge con molte complicazioni[778], e che strumenti come *chatbot* e assistenti vocali sono generalmente sviluppati proprio per rendere più agevole l'utilizzo di siti web e portali percepiti come poco accessibili. L'intelligenza artificiale, dunque, interverrebbe in tali contesti anche al fine di mitigare le possibli diseguaglianze generate dalla digitalizzazione. D'altro canto, però, non può negarsi che si siano già verificate situazioni in cui l'informatizzazione di alcune attività della pubblica amministrazione è sembrata scontrarsi con le caratteristiche di determinate fasce di utenti. Nel nostro paese, ad esempio, la digitalizzazione delle modalità di richiesta e accesso di determinati servizi rivolti principalmente alla terza età ha dato, talvolta, adito a polemiche che è difficile non considerare, almeno in parte, giustificate[779]. I potenziali effetti discriminatori, e le condizioni sociali, economiche e culturali medie dei destinatari, dunque, dovrebbero essere sempre tenuti in considerazione al momento di informatizzare attività della pubblica amministrazione basate sull'interazione col pubblico e in precedenza svolte con modalità in tutto o in parte tradizionali. Le opportunità offerte dall'intelligenza artificiale confermano e

[777] Secondo il più recente aggiornamento dell'indice DESI (*Digital Economy and Society Index*), infatti, l'Italia risulta ad 18° posto in quanto a digitalizzazione della società, nonostante, dopo l'uscita del Regno Unito dall'Unione, essa sia la terza economia per dimensioni. Cfr. EUROPEAN COMMISSION, *DESI country overview – Italy*, https://digital-strategy.ec.europa.eu/en/policies/desi, 2022 (10 agosto 2022). Il giudizio del *country report* è particolarmente severo proprio dal punto di vista delle competenze digitali della popolazione generale: «dagli indicatori di quest'anno emerge che l'Italia sta colmando il divario rispetto all'Unione europea in fatto di competenze digitali di base; ancor oggi però oltre la metà dei cittadini italiani non dispone neppure di competenze digitali di base. La percentuale degli specialisti digitali nella forza lavoro italiana è inferiore alla media dell'UE e le prospettive per il futuro sono indebolite dai modesti tassi di iscrizione e laurea nel settore delle TIC. Se si desidera che l'UE consegua l'obiettivo del decennio digitale in termini di competenze digitali di base e specialisti TIC, è assolutamente necessario un deciso cambio di passo nella preparazione dell'Italia in materia di competenze digitali» (p. 3-4). Si rimanda, per approfondimenti, al report specifico in materia di *human capital* digitale, disponibile alla stessa pagina.

[778] Cfr., tra i molti riferimenti possibili, J. REDMAN; D. R. FLETCHER, *Violent bureaucracy: A critical analysis of the British public employment service*, in *Critical Social Policy*, 42, 2, 2022; A. SARAT, *". . . The law is all over": power, resistance and the legal consciousness of the welfare poor*, in P. EWICK (A CURA DI),*Consciousness and deology*, Londra, 2006 p. 37 ss.; L. GUI, *L'utente che non c'è. Emarginazione grave, persone senza dimora e servizi sociali*, Milano, 1995; E. BOTTA; I. STANZIONE, *Una ricerca esplorativa sulla condizione degli studenti stranieri all'ultimo anno del percorso di studi alla Sapienza Università di Roma*, in *Formazione, lavoro, persona*, 36, 2022, p. 150-174.

[779] Si pensi, ad esempio, alle polemiche generate dalla difficoltà di molti anziani ad accedere al Sistema Pubblico di Identità Digitale e ad altri servizi online, cfr. A. LANA, *Anziani e tecnologia: dallo Spid ai servizi bancari un divario sempre più incolmabil*e, Corriere della Sera, 1 aprile 2022; F. FORMICA, *Odissea Spid: tre soluzioni per ottenerlo in modo rapido*, la Repubblica, 26 marzo 2022. L'impatto del *digital divide*, non solo nella società italiana, è stato messo in luce anche dalla pandemia di Covid-19, nel corso della quale migliaia di studenti hanno avuto forti difficoltà ad accedere alla didattica a distanza, cfr. F. FLORIO, *Torna la didattica a distanza, ma il digital divide taglia fuori 300 mila studenti: «Rischio crisi educativa»,* Open, 3 novembre 2020; *Didattica a distanza, il digital divide si registra in tutti i paesi*, Orizzonte scuola, 31 maggio 2020.

rafforzano questa constatazione: in primo luogo, perché permettono di automatizzare procedure molto complesse; in secondo luogo, perché il loro impiego con funzioni di assistenza all'utente potrebbe far percepire come in tutto o in parte risolti i problemi connessi alla mancanza di competenze digitali in alcune fasce della popolazione, quando in realtà tali competenze sono, nella maggior parte dei casi, necessarie anche per interagire con tali sistemi[780].

Non si può nascondere, inoltre, come la digitalizzazione abbia, spesso, anche finalità di razionalizzazione della spesa pubblica. Da questo punto di vista, l'impiego di *chatbot*, anche relativamente semplici come quelli menzionati, permette evidenti economie di scala: un solo programma informatico interattivo sostituisce, in tutto o in parte, l'attività di un numero indeterminato di operatori e consulenti in carne ed ossa. Come già detto, queste tecnologie incontrano spesso il favore degli utenti, e hanno un dimostrato effetto facilitatore dell'utilizzo che questi fanno delle piattaforme della pubblica amministrazione. La situazione, però, non è priva di ombre:*chatbot*e assistenti vocali sono efficaci solo per funzioni standardizzate, o che, comunque, si presentano con una frequenza statisticamente significativa. Per quanto i progressi nel campo siano rapidi, e tali strumenti sempre più versatili, allo stato dell'arte anche le tecnologie più complesse non sono in grado di rispondere ad ogni richiesta dell'utente umano (se ciò avvenisse, dovremmo considerare superato il test di Turing). Circostanze individuali non comuni o richieste di eccezioni potrebbero risultare oltre le capacità di funzionamento anche dei sistemi di supporto automatico più avanzati (tacendo la circostanza che non appare sempre realistico, viste le risorse limitate a disposizione, che le amministrazioni si dotino degli strumenti migliori). Dovrebbe essere, quindi, sempre possibile sopperire a tali limiti di funzionamento con l'intervento di un operatore umano o utilizzando un canale di comunicazione non digitale, al fine di non lasciare senza risposta istanze potenzialmente anche molto delicate, qualora siano coinvolti servizi essenziali. Eventuali obiettivi di riduzione e ottimizzazione della spesa pubblica non dovrebbero mai portare a ridurre il margine di tutela di diritti primari. Il problema non pare da sottovalutare: la difficoltà ad ottenere assistenza da parte di un essere umano al momento di interagire con grandi operatori privati che hanno automatizzato, per ragioni prima di tutto economiche, i propri sistemi di assistenza ai clienti (è il caso di diverse aerolinee, o di alcune compagnie di telefonia mobile) appartiene già all'esperienza quotidiana di molti di noi[781]. Evitare che la pubblica amministrazione segua la stessa strada,

[780]La necessità di una riflessione sugli effetti in materia di *digital divide* della crescente diffusione di *chatbot* è sottolineata anche da A. FØLSTAD, P. B. BRANDTZÆG, *Chatbots and the new world of HC*, in *Interactions*, 24, 4, 2017, p. 41-42.

[781] Cfr.A. GLASER, *When Robots Make Us Angry, Humans Pay the Price*, Slate, 14settembre 2017, https://bit.ly/3BBQwUB (13 agosto 2022); P. PAIKENS, A. ZNOTIŅŠ, G. BĀRZDIŅŠ, *Human-in-the-loop conversation agent for customer service*, in E. MÉTAIS, F. MEZIANE, H. HORACEK, P. CIMIANO (A CURA DI), *Natural Language Processing and information systems*, Cham, 2020; C. R. TAYLOR, P. J. KITCHEN, M. E. SARKEES, C. O. LOLK,

automatizzando in modo pressoché totale le modalità di prima interazione col cittadino, pare fondamentale in ragione del rilievo di molte funzioni che essa svolge. L'automazione dei canali di accesso ordinari a un numero crescente di servizi pubblici, infatti, potrebbe portare, come evidenziato, in mancanza di cautele idonee, all'esclusione di soggetti deboli, perché in situazioni di svantaggio socio-culturale, o gravati da bisogni specifici e poco frequenti (come una condizione di disabilità o una malattia rara). Considerazioni, queste, che fanno intuire come il diritto a uno *human in the loop* possa rappresentare un presidio fondamentale in situazioni molto diverse dai soli processi decisionali della pubblica amministrazione. Del resto, la totale automazione dei mezzi di comunicazione tra cittadino e amministrazione, come appena visto, potrebbe portare alla radicale esclusione da tali processi di soggetti con necessità non comuni, spesso associate alla marginalità, vanificando ogni tutela eventualmente garantita nelle successive fasi del procedimento e al momento della decisione, umana o automatizzata[782].

Anche altre delle applicazioni dell'intelligenza artificiale viste al paragrafo precedente sembrano esporsi a critiche simili. Se l'uso di tecnologie avanzate per raggiungere livelli di efficienza mai ottenuti nell'erogazione di determinati servizi pubblici è senz'altro da incoraggiare (può essere il caso, ad esempio, del visto sistema di gestione dei flussi idrici sperimentato a New York, Baltimora e San Francisco), altri utilizzi dell'intelligenza artificiale devono essere analizzati con maggiore cautela. In particolare, l'elaborazione di modelli predittivi per la gestione di risorse pubbliche con tecniche di intelligenza artificiale sembra poter facilmente celare, nascosto dall'alibi dell'efficienza, il fine di ridurre la spesa connessa a servizi da cui dipende l'effettività di numerosi diritti sociali. Un'eventuale razionalizzazione della spesa ottenuta attraverso la tecnologia è di certo da accogliere con favore, qualora la qualità delle prestazioni pubbliche ad essa connesse resti inalterata, o addirittura migliori, com'è stato il caso del *Cincinnati Fire Department*. Diverso è il caso, però, di modelli di gestione, ottenuti con l'analisi dei dati, che portino al risultato finale, consapevolmente perseguito dalle autorità che li utilizzano, o semplice effetto collaterale della volontà di ottenere maggior efficienza, di privare di servizi individui in condizione di bisogno effettivo[783]. Né si può

Addressing the Janus face of customer service: a typology of new age service failures, in *European journal of marketing*, 54, 10, 2020.

[782] Infatti, tale situazione di sostanziale esclusione dall'accesso a determinati servizi e funzioni della pubblica amministrazione, in ragione della loro totale automazione, non pare potersi ricondurre al diritto a uno *human in the loop*, allo stato dell'arte previsto, nei pochi ordinamenti che lo sanciscono con fonti di *hard-law*, unicamente nel momento decisionale. Ciò è vero, in primo luogo, per l'art. 22 del GDPR, come già visto in più punti del lavoro. In tali casi, infatti, il diniego di una determinata prestazione pubblica non avviene a valle di una decisione, ma a monte, risultando impossibile, per il soggetto interessato, interfacciarsi con l'Amministrazione. Adottando le categorie del diritto amministrativo, il procedimento col quale il cittadino intendeva soddisfare un determinato interesse, nemmeno è avviato. Pare più che dubbio che tale difficoltà a rivolgersi ai poteri pubblici, causata dall'automazione, sia assimilabile a una decisione automatizzata di segno negativo ai sensi dell'art. 22 GDPR.

[783] Non può mai nascondersi, in sostanza, l'importanza della scelta che sottendono, l'insopprimibile contenuto etico di quest'ultima e il suo legame con l'effettività di diversi diritti sociali, o addirittura con la protezione della vita e

trascurare la possibilità che i *dataset* di partenza siano viziati da *bias* di varia natura (magari specchio di discriminazioni esistenti, per ragioni più varie, nell'effettività dell'erogazione di servizi pubblici), col risultato di un impatto sproporzionato su alcuni gruppi sociali[784]. Una considerazione, quest'ultima, da estendere anche alle viste applicazioni dell'analisi dei dati con finalità di monitoraggio e controllo. Tali strumenti, infatti, risultano in molti casi insostituibili per le capacità di analisi di ambiti estremamente ramificati e complessi che garantiscono, impensabili per l'operatore umano, come nel citato esempio dei mercati finanziari[785]. Non può trascurarsi, però, la possibilità di risultati discriminatori, specialmente qualora tali sistemi siano impiegati al fine di orientare l'attività di indagine e sanzionatoria in campi che implicano l'analisi di dati raccolti in contesti estremamente diversificati dal punto di vista geografico, etnico ed economico. Si pensi, per limitarsi a un solo esempio, all'eventuale utlizzo di sistemi di questo genere per il controllo della corretta erogazione di sussidi a sostegno dei redditi più bassi.

È opportuno sottolineare che l'analisi dei dati con strumenti di intelligenza artificiale, al fine di orientare l'azione generale della pubblica amministrazione, come nei casi, appena analizzati, dei sistemi impiegati per una gestione più razionale delle risorse, o per finalità di monitoraggio e controllo, sottende un'attività che è, in senso lato, decisionale. Da questo punto di vista, l'elemento di distinzione con l'automazione di decisioni che riguardino il singolo individuo sta non tanto nell'assenza di una valutazione, ma nella sua generalità[786]. Gli strumenti di intelligenza artificiale, infatti, non sono direttamente impiegati nella decisione di un caso singolo. Ciò può rendere particolarmente difficile contrastarne eventuali effetti avversi, poiché la decisione riguardante l'individuo – per ipotesi, di diniego dell'accesso a un servizio – non proviene direttamente da essi,

---

dell'integrità fisica. Sulla dimensione valoriale, e dunque in senso lato politica, delle scelte allocative di risorse pubbliche, cfr. ad esempio F. SCIACCA, *I diritti sociali alla prova*, in *Rivista di filosofia del diritto*, 1, 2022, p. 182-183 eancoraG. CALABRESI, P. BOBBITT, *Tragic Choices cit.*

[784]*Ex multis*, si rimanda a B. FRIEDMAN, H. NISSENBAUM, *Bias in computer systems cit.*, in *ACM Transactions on Information Systems*, 3, 1996, doi.org/10.1145/230538.230561; R. DOBBE S. DEAN, T. GILBERT, N. KOHLI, *A Broader View on Bias in Automated Decision-Making: Reflecting on Epistemology and Dynamics*, 2018, arXiv:1807.00553 (14 maggio 2022); J. SILBERG, J. MANYIKA., *Notes from the AI frontier: Tackling bias in AI (and in humans)*, in *McKinsey Global Institute*, 2019, https://mck.com/3ih2l6L; D. DANKS, A. J. LONDON, *Algorithmic bias in autonomous systems*, in*Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*, 17, 2017, p. 4691-4697; S. SILVA, M. KENNEY, *Algorithms, Platforms, and Ethnic Bias*, in *Communications of the ACM*, 62, 11,p. 37–39; S. SILVA, M. KENNEY, *Algorithms, Platforms, and Ethnic Bias: An Integrative Essay*, in *Phylon*, 55, 1-2, 2018, p. 9–37; B. FRIEDMAN, H. NISSENBAUM, *Bias in computer systems cit*., già citati nel corso dell'ampia analisi del tema della discriminazione algoritmica svolta *supra*, p. 140 ss.

[785] Cfr. ancora *Rapporto 2/2021 – L'impiego dell'IA nell'attività di Banca d'Italia, cit*.; *Rapporto 1/2021 – L'impiego dell'IA nell'attività di CONSOB, AGCOM e ARERA, cit., 211 ss*.; *A. FERNANDEZ, Artificial Intelligence in Financial Services, cit*.

[786]È doveroso richiamare, per le considerazioni svolte in queste righe, le già citate riflessioni di G. CALABRESI, P. BOBBITT, *Tragic Choices cit.*, che, nell'allocazione di risorse scarse, individuano due scelte: la decisione atomistica sul caso singolo (c.d. *second-order choice*), che individua l'individuo determinato che accederà a un determinato beneficio, e la scelta generale che, a monte, ha generato la condizione di scarsità (c.d. *first-order choice*), il cui contenuto etico-morale, spesso, non è percepito a sufficienza. Cfr. anche supra, p. 237-238 e, in particolare, quanto esposto nell'ultimo capitolo del lavoro, in materia di intelligenza artificiale e scelte tragiche in ambito medico.

ma da un procedimento amministrativo del quale è, spesso, interamente responsabile un funzionario umano. Qualora il ruolo di tali sistemi sia limitato alla definizione delle macropolitiche di ripartizione delle risorse pubbliche, metterne in discussione le ripercussioni sul caso singolo (che attiene non alla ripartizione di tali risorse, ma all'allocazione individuale di una frazione della quota già ripartita) può essere molto complesso, poiché esse si collocano a monte dell'intero processo decisionale. L'eventualità che tali sistemi intelligenti si comportino, in tutto o in parte, come *black-box* – la circostanza non pare da escludere, vista l'efficacia delle reti neurali profonde per l'analisi predittiva di grandi moli di dati –aggiunge, ovviamente, ulteriore complessità.

La circostanza, dunque, non fa che confermare l'importanza del riconoscimento di una soglia minima di spiegabilità e controllo umano quale garanzia generale in materia di tecnologie intelligenti. Lo spazio di applicazione del diritto a una spiegazione dei risultati di un sistema non può circoscriversi – come fanno le normative più avanzate in materia oggi in vigore, analizzate ai capitoli precedenti – al ristretto ambito di applicazione della decisione individuale in tutto o in parte automatizzata. Tale impostazione, come già analizzato, rischia di generare importanti vuoti di tutela, in una grande varietà di casi in cui l'intelligenza artificiale è impiegata in attività non decisionali o per valutazioni non definibili come individuali. L'impiego, appena citato, per la definizione degli orientamenti generali dell'attività amministrativa è solo un possibile esempio.

1.3 Intelligenza artificiale, Pubblica Amministrazione e nuovi diritti in alcuni casi giudiziari di fronte alle corti di Stati Uniti, Italia e Francia

L'uso dell'intelligenza artificiale da parte della pubblica amministrazione è stato, negli ultimi anni, discusso dalle corti di alcuni paesi. Questo paragrafo sarà dedicato all'analisi di quattro di questi primi casi giudiziari, che appaiono particolarmente rilevanti per diverse ragioni, come la particolare sensibilità del contesto di applicazione dell'intelligenza artificiale, il grado dell'organo giudiziario chiamato a pronunciarsi, o la completezza del ragionamento della corte, particolarmente adatto ad essere inquadrato dal punto di vista dei nuovi diritti.

Due delle vicende analizzate sono state discusse in tribunali degli Stati Uniti, l'ordinamento in cui l'utilizzo dell'intelligenza artificiale da parte dei poteri pubblici è finito più di frequente di fronte al giudice. Il particolare attivismo delle corti statunitensi - i casi aventi ad oggetto le tecnologie intelligenti, negli ultimi anni, possono quantificarsi, pur nella difficoltà di ricostruire un quadro unitario, in alcune decine – si spiega facilmente col tasso di digitalizzazione particolarmente elevato della società statunitense rispetto ad altri paesi avanzati (inclusa gran parte dell'Europa)[787]. Tale

[787] Cfr. ad esempio EUROPEAN INVESTMENT BANK, *Who is prepared for the new digital age? - Evidence from the EIB Investment Survey*, 2020, doi:10.2867/03951; P. NOUVEAU, *European Union's digital governance vs. United States' digital dominance*, in *Revue de la Faculté de droit de l'Université de Liège*, 2, 2020, p. 208-232.

spiccata apertura verso la tecnologia è, ovviamente, affiancata da un sistema di garanzie che rende più facile e frequente l'accesso al giudice di cittadini che si ritenessero pregiudicati da determinati utilizzi dell'intelligenza artificiale, rispetto a paesi, in primo luogo la Cina, che investono particolarmente sull'impiego di sistemi avanzati da parte dei poteri pubblici, ma non possono considerarsi democratici[788]. Peraltro, le controversie giunte di fronte alle corti statunitensi, fino ad ora, si sono concluse spesso con accordi transattivi; i casi in cui i giudici americani hanno avuto modo di esprimere compiutamente la loro opinione sulle garanzie che dovrebbero circondare il coinvolgimento dell'intelligenza artificiale nelle attività della pubblica amministrazione si contano sulle dita di una mano[789].

Barry v. *Lyon*, il primo caso che si è scelto di approfondire, è uno degli esempi più risalenti di automazione di determinate decisioni amministrative giunto di fronte alle corti[790]. Walter Barry, in condizioni di povertà e affetto da una disabilità cognitiva, si vede revocare dal Michigan Department of Human Services, il 1 febbraio 2013, l'assegno di supporto alle spese alimentari di cui è titolare[791]. A decidere il suo caso è stato un algoritmo di recente introduzione, che confronta l'elenco degli aventi diritto al beneficio e i dati del casellario giudiziario dello stato: qualora risultino dei precedenti, il sistema sospende in automatico l'erogazione (come dispongono le norme dello stato) e genera una lettera di notifica come quella che gli è stata recapitata. Il procedimento a suo carico riguarda però il fratello, e il sistema glielo attribuisce per un errore nella compilazione del casellario. Barry chiede di essere ascoltato dalle locali autorità di polizia, un diritto della cui esistenza è la stessa missiva ad informarlo. Chiarito l'equivoco, il bonus è ripristinato. Ciò nonostante, pochi mesi dopo, il primo giugno 2013, Walter Barry riceve una nuova comunicazione della sospensione del beneficio.È necessario che compaia di nuovo di fronte alle autorità di polizia per risolvere il problema. Nononostante riceva rassicurazioni specifiche sul punto dal Michigan Department of Human Services, il problema non viene risolto, né il casellario corretto. Il 16 settembre, il supporto alimentare viene sospeso per la terza volta dal sistema automatizzato. È necessaria un'ulteriore udienza per chiarire, ancora una volta, l'equivoco. Accade lo stesso a diversi

[788] Basti pensare che il già citato *New Generation Artificial Intelligence Development Plan*del 2017 fissa, per il paese, l'obiettivo di divenire il *leader* mondiale nel settore dell'intelligenza artificiale entro il 2030. Sulla digitalizzazione del settore pubblico cinese e sul possibile impatto su diritti e libertà, cfr. R. ARCESATI, *E-government and Covid-19: digital china goes global*, in *MERICS*, 17 marzo 2022, https://bit.ly/3CCYi1F (28 luglio 2022); S. MIRACOLA, *How China uses artificial intelligence to control society*, 3 giugno 2019, https://bit.ly/2WJS3Vg (28 luglio 2022).

[789] Cfr. l'*IA Litigation Database* messo a disposizione dalla *George Washington University* di Washington DC, in cui è possibile consultare lo stato di avanzamento di tutti i casi principali riguardanti l'intelligenza artificiale discussi nei tribunali statunitensi: https://bit.ly/3Rbi1t8 (28 luglio 2022).

[790] Cfr. United States District Court for the Eastern District of Michigan, *Walter Barry* v. *Nick Lyon*, 5:13-cv-13185, https://casetext.com/case/barry-v-lyon (28 luglio 2022) e la relativa pronucia d'appello Court of Appeals for the Sixth Cirtcuit, *Walter Barry* v. *Nick Lyon*, 15-1390 (2016), in *Court Listener*, https://www.courtlistener.com/opinion/4251251/walter-barry-v-nick-lyon/ (28 luglio 2022).

[791]Per la ricostruzione dei fatti nel dettaglio cfr. Court of Appeals for the Sixth Cirtcuit, *Walter Barry* v. *Nick Lyon*, 15-1390 (2016), p. 4-7.

altri beneficiari, che, con un'azione collettiva, portano di fronte alla corte distrettuale competente le modalità con cui il Michigan Department of Human Services implementa le politiche dello stato in materia di sospensione dai benefici sociali dei soggetti con carichi pendenti. Il tribunale dichiara fondate le loro istanze, con una pronuncia poi confermata in appello. Entrambe le decisioni contengono affermazioni particolarmente significative ai fini di quest'analisi. Le corti censurano, in primo luogo, la totale automazione della decisione, che dovrebbe invece prevedere una revisione umana prima dell'effettiva sospensione del beneficio. La pronuncia d'appello, in particolare, osserva: «although the state statute directs MDHHS to "take all reasonable and necessary mesaures using the avalilable technology to ensure the accuracy of this comparison before notifyng a local office of astanding felony warrant or extradition warrant" (Mich. Comp. LAws par. 400.10c.1), there is no directive to determine whether the putative felon is being "actively sought" for prosecution before his or her name is deleted automatically from the list of eligible recipients»[792]. In secondo luogo, i giudici evidenziano la totale inadeguatezza della lettera di notifica generata automaticamente dal sistema, che fornisce all'individuo pregiudicato informazioni generiche e insufficienti. L'arresto di primo grado, sul punto, rileva, con considerazioni poi riportate anche nella sentenza d'appello: «The notice states the intended action – denail or reduction of benefits […]. The notice shows the relevant action as "closed" and states: "your ongoing benefit has been cancelled, but you will continue to recieve benefits through the day before the period listed above". The reason given for the action is simply "you or a member of your group is not eligible for assistance due to a criminal justice disqualification". This fail to indicate whose conduct is at issue. It also fails to indicate which of the five types of criminal justice disqualifications applied by Michigan Department of Human Services is being invoked […] The recipient thus has no basis for making an informed decision whether to contest the disqualification, nor what issues need to be addressed at a hearing»[793].

Il caso *Barry* è significativo, in primo luogo, per essere relativamente risalente nel tempo. I fatti risalgono al 2013, la pronuncia di primo grado all'anno successivo, il processo d'appello si è chiuso nel 2016. Ciò nonostante, le decisioni dei giudici americani a riguardo contengono già molte delle considerazioni che, negli anni successivi, appariranno sempre più urgenti, e che portano, quasi un decennio dopo, a qualificare, in questo lavoro, alcune garanzie che devono circondare l'automazione come nuovi diritti fondamentali. Le sentenze del caso *Barry v. Lyon*, infatti, identificano due grandi criticità del sistema: la totale automazione della procedura di revoca del beneficio sociale in questione e l'insufficienza delle informazioni fornite al soggetto interessato. Le

[792]Court of Appeals for the Sixth Cirtcuit, *Walter Barry* v. *Nick Lyon*, 15-1390 (2016), p. 13.
[793]*Ibidem*, p. 15.

parole della pronuncia d'appello riportate poche righe sopra evidenziano che il quadro normativo volto a regolare l'introduzione della tecnologia automatizzata non prevede controlli al fine di «determine whether the putative felon is being "actively sought" for prosecution before his or her name is deleted automatically from the list of eligible recipients». Non è prevista, come invece dovrebbe, una forma di sorveglianza sulla decisione individuale, prima che questa acquisti efficacia e dispieghi i suoi effetti potenzialmente lesivi della dignità di soggetti deboli. A mancare, in poche parole, è uno *human in the loop*, come diverrà più comune riferirsi a tale garanzia negli anni successivi. Le informazioni fornite ai destinatari della decisione, inoltre, sono insufficienti, non permettendo loro di comprendere le reali ragioni della revoca del beneficio di cui erano destinataried, eventualmente, contestare efficacemente tale decisione. Nel caso specifico, la circostanza non è connessa alla spiegabilità della tecnologia in questione, che non pare particolarmente complessa – automatizzava, semplicemente, il confronto di due banche dati – e sulle cui caratteristiche non si dilungano né la pronuncia di primo grado, né quella d'appello. È interessante notare, però, che la lettera contenente tali informazioni era anch'essa prodotta dal sistema, che risultava, quindi, disegnato in modo tale da non fornire alla persona interessata le motivazioni cui avrebbe avuto diritto. La vicenda, dunque, sembra anticipare un tema che, come già ampiamente analizzato, è oggi giustamente al centro dell'attenzione: la diminuzione delle possibilità di comprensione delle ragioni di una decisione, anche amministrativa, a causa del coinvolgimento della tecnologia nel procedimento con cui viene formulata. Nell'ottica adottata da questo lavoro, pare opportuno aggiungere che le lettere di notifica recapitate a Walter Barry e agli altri soggetti interessati mancavano di un'ulteriore informazione fondamentale, trascurata dai giudici che le hanno esaminate, plausibilmente a causa della minor consapevolezza, all'epoca in cui si sono celebrati i processi in esame, della varietà di questioni connesse alla diffusione di tecnologie avanzate. Esse, infatti, non contenevano alcun riferimento alla circostanza che l'intera procedura si era svolto in modo totalmente automatizzato. Una mancanza di cui, ragionando a contrario, è agevole comprendere la gravità: è plausibile, infatti, che, anche in assenza di una motivazione dettagliata, la consapevolezza della delega totale alla tecnologia della decisione avrebbe portato i soggetti pregiudicati ad identificare immediatamente in un malfunzionamento, o nell'assenza di una revisione umana, la causa di eventuali errori nei dati del casellario. Il possibile impatto della decisione in esame sulle stesse possibilità di sopravvivere dignitosamente dei suoi destinatari, riguardando il supporto all'acquisto di beni alimentari, è perfetto esempio delle ragioni che hanno portato, in questo studio, a identificare nel diritto a essere sempre informati dell'esistenza di un sistema intelligente il primo è più basilare dei nuovi diritti fondamentali dell'era dell'intelligenza artificiale.

Il secondo caso statunitense che si è scelto di esaminare, *Arkansas Department of Human Services v. Legerwood et al.*[794], di pochi anni successivo a *Barry v. Lyon*, è paradigmatico delle possibili conseguenze per i diritti fondamentali dell'utilizzo di applicazioni dell'intelligenza artificiale basate sull'analisi dei dati con finalità, dichiarate o meno, di razionalizzazione e diminuzione della spesa pubblica. La vicenda prende le mosse nel 2015, con la decisione, da parte dei servizi sociali dello stato americano dell'Arkansas, di decidere le ore di assistenza domiciliare garantite a pazienti affetti da varie forme di disabilità per mezzo di «a set of complex computer algorithms»[795], con cui analizzare le risposte date da questi ultimi a 286 domande di autovalutazione delle loro condizioni. Fino ad allora, le modalità di assistenza erano determinate discrezionalmente dal personale sanitario incaricato di tali prestazioni, fino a un massimo di 81 ore a settimana. L'entrata in gioco dell'algoritmo determina un aumento dell'assistenza per il 43% degli aventi diritto, una diminuzione per il 47% e il mantenimento del regime precedente per il restante10%. A subire la diminuzione più drastica sono, in molti casi, i pazienti che si trovano nelle condizioni più gravi, diversi dei quali, con un'azione collettiva, citano in giudizio le competenti autorità amministrative. Chiedono il ripristino, con provvedimento d'urgenza, delle condizioni stabilite da infermieri e medici umani, e lamentano di subire, a causa dell'assistenza ridotta che ricevono, pregiudizi non lievi al loro diritto alla salute e alla loro dignità, come mancanza di cibo, impossibilità di lavarsi e cambiarsi per giorni, riduzione della frequenza di terapie fondamentali, isolamento, solitudine e maggior rischio di infortuni, deterioramento globale del quadro clinico di diversi di essi[796].

La corte competente riconosce le loro ragioni e intima ai servizi sociali dello stato dell'Arkansas di interrompere immediatamente l'utilizzo dell'algoritmo, con un provvedimento poi confermato dalla corte d'appello dello stato, in seguito all'impugnazione da parte di questi ultimi. L'ingiunzione diviene permanente el'*Arkansas Department of Human Services* cessa di usare l'algoritmo, peraltro

[794] Cfr. Pulaski County Circuit Court, Fifth Division, *Arkansas Department of Human Services v. Bradley Ledgerwood et al.*, 60CV-17-442 e la relativa pronuncia di secondo grado Supreme Court of Arkansas, *Arkansas Department of Human Services v. Bradley Ledgerwood et. al.*, 530 S.W.3d 336 (2017), in *Justitia US Law*, https://law.justia.com/cases/arkansas/supreme-court/2017/cv-17-183.html (29 luglio 2022).

[795] È il provvedimento di secondo grado Supreme Court of Arkansas, *Arkansas Department of Human Services v. Bradley Ledgerwood et. al.*, 530 S.W.3d 336 (2017) a usare questa espressione (p. 2); si rinvia alla medesima sentenza (p. 2-7) per una ricostruzione dei fatti più approfondita di quella qui proposta. Per un commento della vicenda, oltre che alcune riflessioni generali sulle possibili conseguenze dell'utilizzo di algoritmi per la gestione di servizi di *welfare* così delicati, cfr. L. X. Z. BROWN, M. RICHARDSON, R. SHETTY, A. CRAWFORD, T. HOAGLAND, *Challenging the Use of Algorithm-driven Decision-making in Benefits Determinations Affecting People with Disabilities*, Center for Democracy & Technology – Report, Oct. 2020, https://bit.ly/3R3fyl4 (29 luglio 2022).

[796] «Appellees claimed that they (1) had been forced to go without food, (2) remained in soiled clothes or have gone without bathing, (3) missed key exercises, treatments, or turnings, (4) faced an increased risk of falling, (5) have become more isolated in their homes; (6) have suffered worsened medical conditions directly due to a lack of care; and (7) have considered moving to nursing homes», Supreme Court of Arkansas, *Arkansas Department of Human Services v. Bradley Ledgerwood et. al.*, 530 S.W.3d 336 (2017), p. 5.

non prima di un ulteriore appello alla Corte Suprema dello stato, dicharato *moot*[797]. Le decisioni emesse in via d'urgenza si fondano, oltre che sull'ovvia necessità di metter fine alle gravi conseguenze generate dal sistema, sulla totale mancanza di trasparenza da parte delle autorità dell'Arkansas nell'informare gli utenti dell'automatizzazione del servizio. La comunicazione in proposito fornita ai beneficiari dell'assistenza domiciliare, infatti, non menziona il coinvolgimento di tecnologie in precedenza non utilizzate, non ne spiega caratteristiche e modalità di funzionamento, non motiva il cambiamento di cui dà conto e, anzi, ne nega gli effetti.La pronuncia di secondo grado ne riporta il testo: «Effective January 1, 2016, the ElderChoices waiver program will be renamed ARChoices in Homecare. Your services and provider will remain the same way. […] For now, just know that the name of your current service is being changed to ARChoices in Homecare [la denominazione dell'algoritmo in esame, *ndr*]. You will continue to receive the same services and the DAAS Nurse will provide more information at your next reassessment»[798].Un approccio opposto, in poche parole, a quello auspicato in questo lavoro. Le conseguenze per la salute di molti pazienti causate dalla delega all'algoritmo di decisioni di tal genere sono un altro esempio della necessità di assicurare che chi si interfacci con tecnologie avanzate sia reso consapevole di questa circostanza, riceva una spiegazione dei loro risultati, e non sia esposto ai possibili effetti indesiderati di un sistema lasciato libero di operare al di fuori di ogni controllo umano. È ragionevole presumere, infatti, che se la comunicazione rivolta agli utenti fosse stata ispirata ai nuovi diritti qui in esame, in primo luogo informandoli dell'automazione delle decisioni che li riguardavano, essi avrebbero potuto contestarla con maggior consapevolezza e prima che si verificassero le spiacevoli conseguenze su cui si è poi fondata la loro class action.

I casi statunitensi *Barry v. Lyon* e *Arkansas Department of Human Services v. Legerwood et al.*, come detto, sono particolarmente significativi per l'impatto che le tecnologie coinvolte avevano su alcuni diritti sociali primari e perché, specialmente nel primo caso, rappresentano una tra le prime occasioni in cui i giudici di un paese democratico si sono confrontati con l'automazione della decisione amministrativa. I provvedimenti giudiziali cui hanno dato origine, per quanto di certo avveniristici, dimostrano, comunque, una consapevolezza solo parziale, e almeno in parte ancora acerba, della portata delle conseguenze, anche giuridiche, derivanti dall'avvento dell'intelligenza artificiale. Una visione molto più matura - come detto nella parte precedente del lavoro, forse la più matura del panorama giuridico globale, allo stato dell'arte – emerge da alcune pronunce del Consiglio di Stato italiano, già citate analizzando il tema della discriminazione algoritmica. Si tratta,

[797] A riportare gli eventi è la stessa dichiarazione di *mootness* Supreme Court of Arkansas, *Arkansas Department of Human Services v. Bradley Ledgerwood et. al.*, CV-18-639 (2019), in *Justitia US Law*, https://law.justia.com/cases/arkansas/supreme-court/2019/cv-18-639.html.

[798]Supreme Court of Arkansas, *Arkansas Department of Human Services v. Bradley Ledgerwood et. al.*, 530 S.W.3d 336 (2017), p. 3.

in particolare, delle sentenze n. 2270 e 8474 del 2019 e n. 881 del 2020[799], relative al c.d. caso *buona scuola*.

I fatti all'origine dei tre provvedimenti sono già stati riassunti: la Legge 107/2015 dispone un piano di assuzioni straordinario nel sistema di pubblica istruzione; l'assegnazione degli insegnanti alle rispettive sedi di lavoro in funzione del collocamento nella graduatoria concorsuale e delle loro preferenze avviene con un algoritmo; il risultato finale appare a molti irrazionale e discriminatorio: docenti posizionatisi tra i primi posti si vedono assegnare sedi di lavoro distanti dalla provincia di preferenza, mentre candidati che avevano ottenuto un punteggio inferiore ottengono la cattedra indicata come prima scelta. Ne scaturiscono diversi ricorsi amministrativi al T.A.R. competente, da parte di docenti che si ritengono pregiudicati e contestano, da vari punti di vista, l'impiego dell'algoritmo in questione. L'impugnazione di alcune di tali sentenze di primo grado porta a diverse pronunce del Consiglio di Stato, tra cui quelle in esame. La posizione espressa dai tribunali amministrativi di primo grado era stata di tendenziale chiusura all'utilizzo di algoritmi da parte della pubblica amministrazione, considerandolo incompatibile con le garanzie procedimentali previste dalla L. 241/1990. Appariva significativa, da questo punto di vista, in particolare una sentenza del T.A.R. Lazio del 2018, che aveva affermato: «un algoritmo, quantunque, preimpostato in guisa da tener conto di posizioni personali, di titoli e punteggi, giammai può assicurare la salvaguardia delle guarentigie procedimentali che gli artt. 2, 6, 7, 8, 9, 10 della legge 7.8.1990 n. 241 hanno apprestato»[800]. A detta della pronuncia, inoltre, sarebbe «mancata nella fattispecie una vera e propria attività amministrativa», non potendosi considerare tale l'attività automatizzata svolta dall'algoritmo.

Il Consiglio di Stato, nelle tre pronunce citate, ha rovesciato radicalmente questa impostazione, pur confermando che, nel caso specifico, l'algoritmo del caso *buona scuola* era stato impiegato in modo illegittimo. Il primo, decisivo passo è stato compiuto dalla sentenza n. 2270 del 2019 che ha rigettato la vista tesi del T.A.R. Lazio sulla generale illiceità del provvedimento amministrativo algoritmico. Al contrario, il massimo organo di giustizia amministrativa ha stabilito che l'impiego

[799] Rispettivamente, Cons. St. sez. VI 8 aprile 2019, n. 2270; Cons. St. sez. VI, 13 dicembre 2019, n. 8474; Cons. St. VI, 4 febbraio 2020, n. 881. Per dei commenti in dottrina si rimanda, tra i molti, a B. MARCHETTI, *La garanzia dello "human in the loop" alla prova della decisione amministrativa algoritmica*, in *BioLaw Journal - Rivista di BioDiritto*, 2021, 2, pp. 367-385; G. GALLONE, A. G. OROFINO, *L'intelligenza artificiale al servizio delle funzioni amministrative: profili problematici e spunti di riflessione. Nota a sent. Cons. Stato sez. VI 4 febbraio 2020 n. 881*,in *Giurisprudenza italiana*, 2020, 7, p. 1738-1748; N. Muciaccia, *Algoritmi e procedimento decisionale: alcuni recenti arresti della giustizia amministrativa*, in *federalismi.it*, 2020, 10, p. 344-368, oltre ai già citati A. NICOTRA, V. VARONE, *L'algoritmo, intelligente ma non troppo*, in *Rivista AIC*, 4, 2019, p. 86-106; E. COCCHIARA, *Procedimento amministrativo e "buon algoritmo"*, in *amministrativ@mente*, 3, 2020, p. 370-385; L. MUSSELLI, *La decisione amministrativa nell'età degli algoritmi: primi spunti*, in *MediaLaws –Rivista di diritto dei media*, 1, 2020, p. 18-28; A. SIMONCINI, *L'algoritmo incostituzionale: l'intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019, p. 63 ss.

[800]T.A.R. Lazio sez. IIIbis,10 settembre 2018, n. 9224.

di tecnologie avanzate per rendere più rapida, precisa ed efficace l'attività amministrativa è non solo legittimo, ma doveroso, contribuendo al perseguimento dei canoni di efficienza, economicità e buon andamento della pubblica amministrazione[801]. Allo stesso tempo, la pronuncia ha chiarito che l'avvento degli algoritmi non può portare a derogare i principi generali che governano l'azione amministrativa. Il Consiglio di Stato ha definito l'algoritmo una "formula tecnica" cui corrisponde una normale "regola giuridica", soggetta ad ogni requisito e garanzia previsti per l'atto amministrativo. Decisivo, in particolare, è che la «decisione robotizzata […] sia conoscibile, secondo una declinazione rafforzata del principio di trasparenza, che implica anche quello della piena conoscibilità di una regola espressa in un linguaggio differente da quello giuridico»[802]. Da questo punto di vista, la sentenza ha chiarito che la tecnologia utilizzata deve risultare comprensibile in ogni componente: «dai suoi autori, al procedimento usato per la sua elaborazione, al meccanismo di decisione, comprensivo delle priorità assegnate nella procedura valutativa e decisionale e dei dati selezionati come rilevanti. Ciò al fine di poter verificare che gli esiti del procedimento robotizzato siano conformi alle prescrizioni e alle finalità stabilite dalla legge o dalla stessa amministrazione a monte di tale procedimento e affinché siano chiare – e conseguentemente sindacabili – le modalità e le regole in base alle quali esso è stato impostato»[803].

È immediato accostare queste parole del Consiglio di Stato a quanto già analizzato in materia di diritto a una spiegazione. Il collegamento coi nuovi diritti analizzati in questo lavoro risulta, in ogni caso, ancora più evidente analizzando il passo successivo compiuto dall'organo di vertice della giustizia amministrativa italiana, nelle già citate sentenze n. 8474 del 2019 e 881 del 2020.In tali sentenze, il Consiglio di Stato, in primo luogo,ha confermato l'auspicabilità di un sempre maggiore coinvolgimento di tecnologie avanzate nell'attività amministrativa, anche discrezionale. In secondo luogo, ha identificato alcuni principi-guida a cui deve ispirarsi la loro implementazione. Lepronunce hanno individuato, in particolare, due esigenze fondamentali: «a) la piena conoscibilità a monte del modulo utilizzato e dei criteri applicati; b) l'imputabilità della decisione all'organo titolare del potere, il quale deve poter svolgere la necessaria verifica di logicità e legittimità della scelta e degli

[801] «Per quanto attiene più strettamente all'oggetto del presente giudizio, devono sottolinearsi gli indiscutibili vantaggi derivanti dalla automazione del processo decisionale dell'amministrazione mediante l'utilizzo di una procedura digitale ed attraverso un "algoritmo" – ovvero di unasequenza ordinata di operazioni di calcolo – che in via informatica sia in grado di valutare e graduare una moltitudine di domande.L'utilità di tale modalità operativa di gestione dell'interesse pubblico è particolarmente evidente con riferimento a procedure seriali o standardizzate, implicanti l'elaborazione di ingenti quantità di istanze e caratterizzate dall'acquisizione di dati certi ed oggettivamente comprovabili e dall'assenza di ogni apprezzamento discrezionale. Ciò è, invero, conforme ai canoni di efficienza ed economicità dell'azione amministrativa (art. 1 l. 241/90), i quali, secondo il principio costituzionale di buon andamento dell'azione amministrativa (art. 97 Cost.), impongono all'amministrazione il conseguimento dei propri fini con il minor dispendio di mezzi e risorse e attraverso lo snellimento e l'accelerazione dell'iter procedimentale», cfr. Cons. St. sez. VI 8 aprile 2019, n. 2270, par. 8.1.

[802] *Ibidem*, par. 8.3.

[803] *Id.*

esiti affidati all'algoritmo».[804] La definizione di conoscibilità adottata è quella, completa, già enunciata nella precedente sentenza 2270 del 2019, che il Consiglio di Stato, nei due provvedimenti, ha collegato agli artt. 13, 14, 15 del GDPR, inclusi, nell'analisi condotta nei paragrafi precedenti, tra i principali indicatori di un crescente riconoscimento giuridico del diritto alla spiegazione[805]. Ambo le pronunce, peraltro, hanno esplicitato che dev'essere garantito a ogni individuo, prima di tutto, il «diritto a conoscere l'esistenza di processi decisionali automatizzati che lo riguardino». In materia di imputabilità della decisione all'organo titolare del corrispondente potere amministrativo, invece, il Consiglio di Stato ha identificato il «principio di non esclusività della decisione algoritmica», facendo riferimento, anche in questo caso, al GDPR, e in particolare al più volte menzionato art. 22[806]. I giudici amministrativi, ancora una volta con la medesima espressione in entrambe le pronunce, ne hanno dato una definizione molto completa, che non pare necessitare commenti: «In proposito, deve comunque esistere nel processo decisionale un contributo umano capace di controllare, validare ovvero smentire la decisione automatica. In ambito matematico ed informativo il modello viene definito come HITL (human in the loop), in cui, per produrre il suo risultato è necessario che la macchina interagisca con l'essere umano».A completare il quadro di garanzie che, secondo le due pronunce, devono accompagnare l'utilizzo di tecnologie intelligenti nell'attività amministrativa èil «principio di non discriminazione algoritmica», fatto risalire al Cons. 71 del GDPR e già analizzato in profondità nella seconda parte del lavoro[807].

[804] La frase è riferita a ambo le pronunce perché le due sentenze presentano, per lunghi tratti, identica formulazione; cfr., in questo caso, e Cons. St. sez. VI, 13 dicembre 2019, n. 8474, par. 12 e Cons. St. VI, 4 febbraio 2020, n. 881, par. 9.

[805] «In particolare, in maniera innovativa rispetto al passato, gli articoli 13 e 14 del Regolamento stabiliscono che nell'informativa rivolta all'interessato venga data notizia dell'eventuale esecuzione di un processo decisionale automatizzato, sia che la raccolta dei dati venga effettuata direttamente presso l'interessato sia che venga compiuta in via indiretta. Una garanzia di particolare rilievo viene riconosciuta allorché il processo sia interamente automatizzato essendo richiesto, almeno in simili ipotesi, che il titolare debba fornire "informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento per l'interessato" . In questo senso, in dottrina è stato fatto notare come il legislatore europeo abbia inteso rafforzare il principio di trasparenza che trova centrale importanza all'interno del Regolamento. L'interesse conoscitivo della persona è ulteriormente tutelato dal diritto di accesso riconosciuto dall'articolo 15 del Regolamento che contempla, a sua volta, la possibilità di ricevere informazioni relative all'esistenza di eventuali processi decisionali automatizzati», cfr. Cons. St. sez. VI, 13 dicembre 2019, n. 8474 par. 13 e Cons. St. VI, 4 febbraio 2020, n. 881, par. 10.

[806] «In secondo luogo, l'altro principio del diritto europeo rilevante in materia (ma di rilievo anche globale in quanto ad esempio utilizzato nella nota decisione Loomis vs. Wisconsin), è definibile come il principio di non esclusività della decisione algoritmica. Nel caso in cui una decisione automatizzata "produca effetti giuridici che riguardano o che incidano significativamente su una persona", questa ha diritto a che tale decisione non sia basata unicamente su tale processo automatizzato (art. 22 Reg.). In proposito, deve comunque esistere nel processo decisionale un contributo umano capace di controllare, validare ovvero smentire la decisione automatica. In ambito matematico ed informativo il modello viene definito come HITL (human in the loop), in cui, per produrre il suo risultato è necessario che la macchina interagisca con l'essere umano», cfr. Cons. St. sez. VI, 13 dicembre 2019, n. 8474 par. 15.2 e Cons. St. VI, 4 febbraio 2020, n. 881, par. 11.2.

[807] «In terzo luogo, dal considerando n. 71 del Regolamento 679/2016 il diritto europeo trae un ulteriore principio fondamentale, di non discriminazione algoritmica, secondo cui è opportuno che il titolare del trattamento utilizzi procedure matematiche o statistiche appropriate per la profilazione, mettendo in atto misure tecniche e organizzative adeguate al fine di garantire, in particolare, che siano rettificati i fattori che comportano inesattezze dei dati e sia minimizzato il rischio di errori e al fine di garantire la sicurezza dei dati personali, secondo una modalità che tenga

Come già evidenziato in precedenza, i recenti approdi del Consiglio di Stato sembrano, allo stato dell'arte, l'elaborazione giurisprudenziale più ricca, matura e consapevole in materia di intelligenza artificiale. Opportunità e criticità connesse all'utilizzo di algoritmi nei procedimenti decisionali sono inquadrate con precisione, e la soluzione offerta è un'apprezzabile sintesi tra l'esigenza di non rinunciare all'innovazione tecnologica e la necessità di conservare le garanzie che normalmente circondano la relazione tra autorità e consociati. Essa, inoltre, appare particolarmente vicina all'approccio basato sui diritti adottato in questo lavoro. I principi generali che identifica, infatti, corrispondono, in quanto al contenuto, ai nuovi diritti già commentati: l'impiego dell'intelligenza artificiale dev'essere dichiarato a ogni interessato, spiegabile, soggetto al controllo di un essere umano. Inoltre, la scelta di ragionare per principi, senza cercare la soluzione alle questioni sottoposte alla corte solo nella specifica disciplina del procedimento amministrativo, dimostra la consapevolezza, da parte del Consiglio di Stato, della necessità di identificare delle garanzie giuridiche generali che presidino l'avvento dell'intelligenza artificiale, al fine di conservare la centralità dell'essere umano. Un'impostazione condivisa da questo lavoro, del quale le pronunce analizzate costituiscono uno dei principali punti di partenza. Al fine di costruire un impianto dogmatico e normativo solido e pienamente in grado di orientare lo sviluppo tecnologico verso la giusta direzione pare opportuno, però, inquadrare, senza reticenze, tali garanzie come diritti fondamentali. Solo in tal modo, infatti, le tutele individuate anche dal Consiglio di stato saranno messe al riparo da possibili ambiguità in materia di titolarità e azionabilità individuale, e avranno un riconoscimento consono alla loro inerenza ad aspetti chiave della sfera della personalità.

Infine, è doveroso menzionare, per il rango dell'organo giudiziale che l'ha emanata, una pronuncia del *Conseil constitutionnel* francese risalente al 2020, relativa all'utilizzo, nella selezione degli studenti aventi accesso a determinati corsi di educazione superiore, di un algoritmo denominato *Parcoursup*[808]. La questione di costituzionalità verteva, nello specifico, sulla legittimità di alcune limitazioni al diritto di accesso ai documenti amministrativi – il parametro di costituzionalità preso in considerazione era l'art. 15 della *Declaration* del 1789[809] - in tale materia, disposte con una

conto dei potenziali rischi esistenti per gli interessi e i diritti dell'interessato e che impedisca tra l'altro effetti discriminatori nei confronti di persone fisiche sulla base della razza o dell'origine etnica, delle opinioni politiche, della religione o delle convinzioni personali, dell'appartenenza sindacale, dello status genetico, dello stato di salute o dell'orientamento sessuale, ovvero che comportano misure aventi tali effetti. In tale contesto, pur dinanzi ad un algoritmo conoscibile e comprensibile, non costituente l'unica motivazione della decisione, occorre che lo stesso non assuma carattere discriminatorio», Cons. St. sez. VI, 13 dicembre 2019, n. 8474, par. 15.3 e Cons. St. VI, 4 febbraio 2020, n. 881, par. 11.3.

[808] *Conseil constitutionnel*, Décision n. 2020-834 QPC du 3 avril 2020, per un commento in dottrina cfr. T. DOUVILLE, *Parcoursup: transparence des algorithmes locaux limitée à raison pour le Conseil constitutionnel: observations sous le Conseil constitutionnel, du 3 avril 2020, n. 2020-834 QPC, UNEF - Qualification de la décision: confirmation*, in *Dalloz IP/IT* , 2020, p.516 ss.

[809] Il quale recita: «La Société a le droit de demander compte à tout Agent public de son administration».

modifica del *Code de l'education* risalente al 2018[810], che aveva in parte derogato alle garanzie introdotte, sul diritto alla spiegazione, dalla *Loi n 2016-1321 du 7 octobre 2016 pour une République numérique* nel *Code des relations entre le public et l'administration*, come visto particolarmente dettagliate ed onerose per la pubblica amministrazione[811]. Nello specifico, tale novella aveva escluso che soggetti terzi potessero accedere alla documentazione riguardante i criteri di funzionamento dell'algoritmo utilizzato, e limitato il diritto d'accesso degli studenti interessati alle sole «informations relatives aux critères et modalités d'examen de leur candidature»[812], una volta che la decisione che li riguardava era stata presa. Il Conseil constitutionnel ha ritenuto conformi a costituzione tali limitazioni, ritenendole necessarie alla tutela dell'interesse concorrente alla segretezza delle deliberazioni delle commissioni incaricate di tali decisioni, e sottolineando che non si trattava, in ogni caso, di procedure interamente automatizzate[813]. Inoltre, il nucleo essenziale del diritto d'accesso era comunque garantito dalla possibilità, per gli studenti interessati, di accedere alle informazioni e alla motivazione della decisione che li riguardava[814]; il Giudice delle Leggi

[810] Loi n. 2018-166 du 8 mars 2018, *relative à l'orientation et à la réussite des étudiants*, il cui art. 1 ha aggiunto all'art. L.612-3 del *Code de l'education*, che disciplina le modalità d'accesso al primo ciclo d'istruzione universitaria, questo *alinea*: «Afin de garantir la nécessaire protection du secret des délibérations des équipes pédagogiques chargées de l'examen des candidatures présentées dans le cadre de la procédure nationale de préinscription prévue au même deuxième alinéa [la procedura di preiscrizione e selezione su base nazionale, che coinvolgeva anche l'impiego dell'algoritmo *Parcoursup*], les obligations résultant des articles L. 311-3-1 et L. 312-1-3 du code des relations entre le public et l'administration sont réputées satisfaites dès lors que les candidats sont informés de la possibilité d'obtenir, s'ils en font la demande, la communication des informations relatives aux critères et modalités d'examen de leurs candidatures ainsi que des motifs pédagogiques qui justifient la décision prise».

[811] Cfr. *supra* p. 205 ss.

[812] Cfr. nota precedente.

[813] «Toutefois, en premier lieu, il ressort des travaux préparatoires que le législateur a considéré que la détermination de ces critères et modalités d'examen des candidatures, lorsqu'ils font l'objet de traitements algorithmiques, n'était pas dissociable de l'appréciation portée sur chaque candidature. Dès lors, en restreignant l'accès aux documents administratifs précisant ces critères et modalités, il a souhaité protéger le secret des délibérations des équipes pédagogiques au sein des établissements. Il a ainsi entendu assurer l'indépendance de ces équipes pédagogiques et l'autorité de leurs décisions. Ce faisant, il a poursuivi un objectif d'intérêt général. En deuxième lieu, la procédure nationale de préinscription instituée à l'article L. 612-3 du code de l'éducation, notamment en ce qu'elle organise les conditions dans lesquelles les établissements examinent les vœux d'inscription des candidats, n'est pas entièrement automatisée. D'une part, l'usage de traitements algorithmiques pour procéder à cet examen n'est qu'une faculté pour les établissements. D'autre part, lorsque ceux-ci y ont recours, la décision prise sur chaque candidature ne peut être exclusivement fondée sur un algorithme. Elle nécessite, au contraire, une appréciation des mérites des candidatures par la commission d'examen des vœux, puis par le chef d'établissement.», *Conseil constitutionnel*, Décision n. 2020-834 QPC du 3 avril 2020, par. 13-14.

[814] La sentenza, infatti, afferma: «En troisième lieu, en application du deuxième alinéa du paragraphe I de l'article L. 612-3, les caractéristiques de chaque formation sont portées à la connaissance des candidats, avant que ceux-ci ne formulent leurs vœux, par l'intermédiaire de la plateforme numérique mise en place dans le cadre de la procédure nationale de préinscription. Elles font l'objet d'un cadrage national fixé par arrêté du ministre de l'enseignement supérieur. Il en résulte, d'une part, que les candidats ont accès aux informations relatives aux connaissances et compétences attendues pour la réussite dans la formation, telles qu'elles sont fixées au niveau national et complétées par chaque établissement. Ils peuvent ainsi être informés des considérations en fonction desquelles les établissements apprécieront leurs candidatures. Il en résulte, d'autre part, que les candidats ont également accès aux critères généraux encadrant l'examen des candidatures par les commissions d'examen des vœux. Si la loi ne prévoit pas un accès spécifique des tiers à ces informations, celles-ci ne sont pas couvertes par le secret. Les documents administratifs relatifs à ces connaissances et compétences attendues et à ces critères généraux peuvent donc être communiqués aux personnes qui en font la demande, dans les conditions de droit commun prévues par le code des relations entre le public et l'administration. En dernier lieu, en application du dernier alinéa du paragraphe I de l'article L. 612-3, une fois qu'une

francese, in ogni caso, ha chiarito che le norme che restringevano le prerogative dei terzi non esentavano, comunque, le commissioni incaricate della selezione dal pubblicare un report riassuntivo della loro attività, una volta terminate le procedure, onde favorire il controllo diffuso sulla loro esattezza e legittimità[815].

Ai fini di questo lavoro, la pronuncia pare rivestire minor interesse rispetto alle altre già analizzate. A venire in gioco, infatti, non è la scarsa comprensibilità dell'algoritmo, per ragioni tecniche o di segreto industriale, o la reticenza nei confronti dei destinatari in merito al suo coinvolgimento nella decisione che li riguarda, ma la legittimità di una scelta puramente normativa di limitare i diritti d'informazione generalmente garantiti nel procedimento amministrativo. A renderne opportuna la trattazione sintetica sono stati, come già detto, il rango dell'organo giudiziario da cui proviene e la circostanza che le norme oggetto del giudizio di costituzionalità chiamassero in causa alcune di quelle, riguardantila decisione automatizzata, largamente analizzate nei capitoli precedenti. Ad ogni modo, nella prospettiva adottata in questo studio la sentenza mette in luce un'interessante circostanza: gli interessati a una spiegazione dei risultati di un sistema possono non consistere in coloro che, nella percezione comune, sembrino subirne gli effetti. È il caso dei potenziali terzi cui la norma discussa dal Conseil constitutional nega l'accesso alle informazioni sulle singole decisioni delle commissioni didattiche, interessati a valutare il funzionamento globale del sistema di selezione universitaria francese[816]. L'estensione dell'eventuale diritto a una spiegazione di questi

---

décision de refus a été prise à leur égard, les candidats peuvent, à leur demande, obtenir la communication par l'établissement des informations relatives aux critères et modalités d'examen de leurs candidatures, ainsi que des motifs pédagogiques justifiant la décision prise à leur égard. Ils peuvent ainsi être informés de la hiérarchisation et de la pondération des différents critères généraux établies par les établissements ainsi que des précisions et compléments apportés à ces critères généraux pour l'examen des vœux d'inscription. La communication prévue par ces dispositions peut, en outre, comporter des informations relatives aux critères utilisés par les traitements algorithmiques éventuellement mis en œuvre par les commissions d'examen», *Conseil constitutionnel*, Décision n. 2020-834 QPC du 3 avril 2020, par. 15-16.

[815] Riguardo a quest'ultimo profilo, la pronuncia è inquadrabile come un'interpretativa di rigetto: «une fois la procédure nationale de préinscription terminée, l'absence d'accès des tiers à toute information relative aux critères et modalités d'examen des candidatures effectivement retenus par les établissements porterait au droit garanti par l'article 15 de la Déclaration de 1789 une atteinte disproportionnée au regard de l'objectif d'intérêt général poursuivi, tiré de la protection du secret des délibérations des équipes pédagogiques. Dès lors, les dispositions contestées ne sauraient, sans méconnaître le droit d'accès aux documents administratifs, être interprétées comme dispensant chaque établissement de publier, à l'issue de la procédure nationale de préinscription et dans le respect de la vie privée des candidats, le cas échéant sous la forme d'un rapport, les critères en fonction desquels les candidatures ont été examinées et précisant, le cas échéant, dans quelle mesure des traitements algorithmiques ont été utilisés pour procéder à cet examen», *Conseil constitutionnel*, Décision n. 2020-834 QPC du 3 avril 2020, par. 17.

[816] Il Conseil constitutionnel, infatti, era stato adito con una *question prioritaire de constitutionnalité* che aveva avuto origine da un ricorso dell'*Union nationale des étudiants de France*, che lamentava l'illegittimità delle norme qui commentate, poiché impedivano l'accesso da parte di terzi alle informazioni riguardante le procedure di preiscrizione e ammissione all'istruzione superiore, anche dal punto di vista delle tecnologie avanzate in esse coinvolte, cfr. *Conseil constitutionnel*, Décision n. 2020-834 QPC du 3 avril 2020, par. 2: «L'union requérante, rejointe par plusieurs parties intervenantes, reproche à ces dispositions de restreindre l'accès aux informations relatives aux critères et aux modalités d'examen, par les établissements d'enseignement supérieur, des demandes d'inscription dans une formation du premier cycle. Selon elle, ces dispositions seraient contraires au droit à la communication des documents administratifs qui découlerait de l'article 15 de la Déclaration des droits de l'homme et du citoyen de 1789. En effet, ces dispositions excluraient tout accès, des candidats comme des tiers, aux algorithmes susceptibles d'être utilisés par les établissements

ultimi, dovrà, allora, misurarsi nel bilanciamento con gli altri interessi in gioco, in primo luogo con le esigenze di riservatezza dei soggetti che abbiano interagito direttamente col sistema. Un bilanciamento del quale la novella del *Code de l'education* introdotta in Francia nel 2018 e la relativa sentenza del Conseil constitutionnel – nonstante il parametro di costituzionalità chiamato in causa, l'art. 15 della *Declaration* del 1789, non possa certo definirsi un nuovo diritto - possono, pur indirettamente, rappresentare un primo esempio.

2. Intelligenza artificiale, giustizia e diritti fondamentali

2.1. L'intelligenza artificiale nel settore della giustizia: una panoramica delle principali applicazioni esistenti e delle possibili prospettive future

L'utilizzo dell'intelligenza artificiale nel settore della giustizia è stato, per ovvie ragioni, estensivamente commentato dalla letteratura giuridica, anche di lingua italiana[817]. Le applicazioni tecnologiche impiegate nel campo sono estremamente diversificate, e interessano pressoché ogni ambito del sistema giustizia inteso in senso ampio: l'attività professionale degli operatori legali, giudiziale e stragiudiziale; l'organizzazione dei tribunali e la gestione delle risorse, materiali e umane, al loro interno; l'automazione, totale o parziale, di segmenti del processo. Il convolgimento di tecnologie intelligenti nell'attività decisionale dei magistrati – sia nella risoluzione definitiva delle controversie, che in altri momenti del procedimento in cui sono chiamati a esprimere una valutazione, come l'adozione di provvedimenti urgenti, o la disposizione di misure cautelari – è stato, ovviamente, oggetto di particolari discussioni[818]. In primo luogo, in ragione del dibatitto che, come già più volte ripetuto, circonda, in generale, l'impiego dell'intelligenza artificiale nei

---

pour traiter les candidatures à l'entrée dans une telle formation, formulées sur la plateforme numérique dite "Parcoursup". Or, une telle exclusion ne serait justifiée ni par le secret des délibérations des jurys ni par aucun autre motif. En outre, ces dispositions méconnaîtraient le droit à un recours juridictionnel effectif, à un double titre. D'une part, elles empêcheraient d'exercer avec succès un recours contre l'absence de communication des informations en cause. D'autre part, elles priveraient les justiciables des éléments nécessaires à la contestation effective du bien-fondé des refus d'inscription».

[817] *Ex multis*, si rimanda ai già citati A. GARAPON – J. LASSÈGUE, *Justice digitale cit.*;A. SANTOSUOSSO, *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritto cit.*; C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi cit.*; S. PENASA, *Intelligenza artificiale e giustizia: il delicato equilibrio tra affidabilità tecnologica e sostenibilità costituzionale cit.*; M. FASAN, *L'intelligenza artificiale nella dimensione giudiziaria cit.*;F. DONATI, *Intelligenza artificiale e giustizia cit.*;U. PAGALLO – S. QUATTROCOLO, *The impact of AI in Criminal Law, and its Twofold Procedures cit.*;S. QUATTROCOLO, *Intelligenza artificiale e giustizia: nella cornice della Carta Etica Europea, gli spunti per un'urgente discussione tra scienze penali e informatiche cit.*; A. ZAVRŠNIK, *Criminal justice, artificial intelligence systems, and human rights cit.*;M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale cit.*;D. POLIDORO, *Tecnologie informatiche e procedimento penale cit.*; W. S. ISAAC, *Hope, Hype e Fear, The Promise and Potential Pitfalls of Artificial Intelligence in Criminal Justice cit.*

[818] Tutti i contributi citati alla nota precedente trattano, in tutto o in parte, il tema specifico dell'intelligenza artificiale nella decisione giudiziaria. Sull'argomento, inoltre, si rimanda, tra gli altri, anche a M. LUCIANI, *La decisione giudiziaria robotica*, in *Rivista AIC*, 3, 2018, https://bit.ly/3f1DdE4 (19 agosto 2022);T. SOURDIN,*Judge v Robot?: Artificial intelligence and judicial decision-making*,in *The University of New South Wales Law Journal*, 4, 41, 2018, p. 1114–1133;F. FAGAN, S. LEVMORE, *The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion*, in *California Law Review*, 1, 2019, p. 1-37; e ai numerosi contributi raccolti in G. SARTOR, K.BRANTING (A CURA DI),*Judicial Applications of Artificial Intelligence*, Dordrecht, 1998.

meccanismi decisionali. In secondo luogo, per la peculiarità della decisione giudiziaria, delegata a un professionista estremamente qualificato e che giunge al termine di un confronto dialogico le cui modalità di svolgimento sono regolate nei dettagli, al fine di garantirne la corretta formazione. Una decisione che disciplina la composizione di diritti e interessi contrapposti o, in ambito penale, definisce i confini della potestà punitiva dello stato, e nei confronti della quale l'individuo è tutelato, nei sistemi democratici, da un corposo insieme di diritti *nel* processo[819]. Non sorprende, allora, che l'attenzione della dottrina si sia concentrata, prima di tutto, sulle applicazioni dell'intelligenza artificiale destinate a supportare, o radicalmente sostituire, il giudicante. Peraltro, la riflessione scientifica ha spesso preso in esame sperimentazioni puntuali o iniziative provenienti dal mondo della ricerca, la cui implementazione effettiva nei sistemi giudiziari appare meramente ipotetica. Le tecnologie più discusse e studiate, allora, spesso non sono quelle col maggior utilizzo pratico. Allo stato dell'arte, le principali applicazioni delle tecnologie intelligenti nel mondo della giustizia, già diffuse sul mercato o da considerarsi ancora semplici ipotesi di lavoro, possono suddividersi nelle seguenti categorie[820]:

- **Sistemi basati sull'elaborazione del linguaggio naturale che facilitano analisi, redazione e revisione di testi giuridici:** uno degli utilizzi dell'intelligenza artificiale oggi maggiormente presente sul mercato, su cui le *law firm* internazionali sembrano disposte a investimenti rilevanti [821]. Si tratta di applicazioni molto varie, che assistono il professionista umano predisponendo bozze di clausole o interi contratti, documenti di altro genere o atti del processo, attraverso tecniche di intelligenza artificiale applicate all'analisi di contratti e documenti già esistenti. Risulta complesso ricostruire il quadro completo dell'uso di questo genere di tecnologie, per ragioni di segreto industriale e

[819] Cfr., tra i molti, le analisi, relative a diversi ordinamenti, nazionali e sovranazionali, di G. VIGLIETTA, *Il giudice penale e i diritti fondamentali*, in *Questione giustizia*, 5, 2007; V. PETRALIA, *Equo processo, giudicato nazionale e convenzione europea dei diritti dell'uomo*, Torino, 2012; R. GOSS, *Criminal fair trail rights: article 6 of the European Convention of Human Rights*, Londra, 2016. Sulle specifiche vicende del c.d. giusto processo nell'ordinamento italiano, v. ad esempio P. FERRUA, *Il giusto processo*, Bologna, 2005; M.L. CAMPIANI, *Giusto processo civile e penale*, Napoli, 2014.

[820] Una panoramica approfondita degli utilizzi dell'intelligenza artificiale nei sistemi giudiziari dei paesi del Consiglio d'Europa è reperibile all'Appendix I della EUROPEAN COMMISSION FO THE EFFICIENCY OF JUSTICE (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment*, Strasburgo, 3-4 dicembre 2018, https://bit.ly/2mkmO2A (19 agosto 2022), p. 13-59. Per un quadro delle principali ipotesi di ricerca, cfr. invece L. F. DE OLIVEIRA ET AL, *Path and future of artificial intelligence in the field of justice: a systematic literature review and a research agenda*, in *SN Social Sciences*, 2, 180, 2022, https://doi.org/10.1007/s43545-022-00482-w (19 agosto 2022).

[821] Cfr. ad esempio A. E. DAVIS, *The Future of Law Firms (and Lawyers) in the Age of Artificial Intelligence*, American Bar Association, 2 ottobre 2020, https://bit.ly/3SbBVF1 (19 agosto 2022); A. HAYAT, *The future of the law: White & Case on artificial intelligence*, Chambers Associate, novembre 2019, https://bit.ly/3BSHgvI (19 agosto 2022); *Law Firms and Their Tech: Clifford Chance, Ashurst, Freshfields, DLA Piper, Mayer Brown*, New Law Academy, agosto 2021, https://bit.ly/3QOznM1 (19 agosto 2022); S. FLYNN, *How Natural Language Processing (NLP) AI Is Used in Law*, Law Technology Today, 9 giugno 2019, https://bit.ly/3eOYH75 (19 agosto 2022). L'importanza di tali strumenti è sottolineata anche dalla COUNCIL OF BARS AND LAW SOCIETIES OF EUROPE, *Guide on the use of artificial intelligence-based tools by lawyers and law firms in the EU*, 2022, https://bit.ly/3dsIkgj (19 agosto 2022), p. 20-38.

perché, più banalmente, i grandi studi professionali non sono tenuti a rivelarne l'utilizzo. Non mancano, comunque, gli esempi di prodotti ormai famosi, nonché le conferme sperimentali della loro efficacia, dimostratasi superiore a quella di operatori umani esperti, ad esempio, nell'individuazione di clausole potenzialmente invalide in contratti complessi[822].

- **Motori di ricerca giurisprudenziali avanzati:** la combinazione di *natural language processing* e apprendimento automatico può portare a raffinare gli attuali motori di ricerca, permettendo di inferire correlazioni più accurate tra risultati e chiavi di ricerca, o similitudini tra sentenze non riscontrabili con l'utilizzo di tecnologie meno complesse[823]. Si tratta, semplicemente, di un'applicazione settoriale dei risultati senza precedenti raggiunti, negli ultimi anni, nell'analisi e traduzione di testi, e analizzati nelle parti precedenti del lavoro[824]. Anche in questo caso, non mancano gli esempi di prodotti che hanno già raggiunto un notevole successo sul mercato grazie all'implementazione di tecniche di intelligenza artificiale, come i motori di ricerca francesi *Doctrine* [825] e *JurisData Analytics*[826] o lo statunitense *Case List Analyzer*[827]. Iniziative imprenditoriali dello stesso tipo sono presenti anche in Italia[828]. L'interesse dell'avvocatura per questi strumenti è senz'altro elevato, mentre il loro utilizzo in ambito pubblico, in primo luogo da parte della magistratura, sembra, in diversi paesi, poter essere frenato dalla mancanza di risorse.
- **Strumenti a supporto dell'organizzazione del lavoro di tribunali, studi professionali e singoli magistrati**: strumenti, in realtà, utilizzabili e utilizzati in ogni organizzazione

[822] Un esaustivo elenco dei prodotti più interessanti è disponibile in *The top players in the AI-powered contract management space*, www.cenza.co, 26 maggio 2022, https://bit.ly/3RYzFBe (19 agosto 2022); cfr. anche COUNCIL OF BARS AND LAW SOCIETIES OF EUROPE, *Guide on the use of artificial intelligence-based tools by lawyers and law firms*, p. 20 ss. Sulle possibilità dischiuse dalle tecnologie avanzate per la redazione di contratti più comprensibili al grande pubblico, cfr. M. C. COMPAGUCCI, M. FENWICK, H. HAAPIO, *Digital technology, future lawyers and the computable contract designer of Tomorrow*, in M. C. COMPAGUCCI, M. FENWICK, H. HAAPIO (A CURA DI), *Research Handbook in Contract Design*, Cheltenham-Northampton, 2022, p. 421-445.

[823] Cfr. sul punto l'analisi della EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment*, p. 16 ss.

[824] Cfr. ad esempio *supra*, p. 47 ss. e, nell'ambito della *content moderation* delle reti sociali, p. 107 ss. In generale, v. ancheS. RUSSELL, P. NORVIG, *Artificial intelligence cit.(4ª ed.)*, p. 26 ss.

[825] Si rimanda al sito del servizio, che presenta *Doctrine* come la prima *plateforme d'intelligence juridique*: https://www.doctrine.fr/ (18 agosto 2022).

[826] Il servizio è offerto da Lexis 360® Intelligence, parte della *branch* francese della multinazionale dei servizi professionali Lexis Nexis, cfr. https://bit.ly/3LnuRTA (19 agosto 2022).

[827] Per maggiori dettagli, cfr. la pagina web del motore di ricerca: https://lexmachina.com/case-list-analyzer/ (19 agosto 2022).

[828]Di particolare interesse sembra, in particolare, un progetto avviato di recente da ONE Legale, società italiana di servizi di informazione per il mercato legale parte del gruppo internazionale Wolters Kluwer, che dal 2021 offre ai propri clienti la *funzionalità Giurimetria*, descritta come un servizio che applica modelli di intelligenza artificiale alla ricerca di giurisprudenza, cfr. *One LEGALE presenta la nuova funzionalità Giurimetria*, Studio Cataldi – il diritto quotidiano, 8 settembre 2021, https://bit.ly/3BRDe6M (19 agosto 2022).

complessa[829]. L'assegnazione delle cause ai magistrati di un ufficio giudiziario o di una sezione di quest'ultimo, o la definizione del calendario delle udienze di un singolo magistrato,potrebbero risultare più razionali ed efficienti se svolte con l'ausilio di strumenti di intelligenza artificiale volti a calcolare la durata prevista e la mole di lavoro probabilmente necessaria per ogni singolo procedimento. Lo stesso, ovviamente, può dirsi della ripartizione del lavoro tra i singoli professionisti di uno studio legale. Si tratta, peraltro, di applicazioni che, in ambito pubblico, rimangono, allo stato dell'arte, al livello di mere ipotesi, e la cui implementazione, in diversi ordinamenti, dovrebbe, probabilmente, essere preceduta da un'apposita modifica delle norme che disciplinano la ripatizione dei carichi di lavoro tra i singoli magistrati di un medesimo ufficio giudiziario[830].

- **Applicazioni dell'intelligenza artificiale a supporto di valutazioni giudiziali già in parte standardizzate**: un utilizzo dell'intelligenza artificiale che rientra, di fatto, tra le ipotesi di automazione totale o parziale della decisione del magistrato, ma che, generalmente, non suscita particolari resistenze nella letteratura di settore[831]. I sistemi giuridici dei paesi avanzati, in primo luogo l'Italia, hanno sviluppato diverse strategie di razionalizzazione di valutazioni giudiziali relative a vicende particolarmente frequenti e che implicano l'attribuzione di un valore economico a determinate situazioni, al fine di limitare, in materia, il divario esistente tra diversi uffici giudiziari o singoli magistrati. Nel nostro ordinamento, ad esempio, la liquidazione del danno non patrimoniale da sinistro stradale, o in seguito a diffazione a mezzo stampa, avvienesulla base di tabelle elaborate dall'*Osservatorio sulla giustizia civile di Milano*[832]. Questi standard, calcolati sulla base dei precedenti stratificati in tali ambiti, potrebbero essere resi più accurati e calzanti al caso concreto qualora tecnologie di intelligenza artificiale – in grado di analizzare un numero maggiore di precedenti, tenendo conto di più variabili – fossero coinvolte nella

[829] Cfr., ad esempio, H. BENBYA, T. H. DAVENPORT, S. PACHIDI, *Artificial Intelligence in Organizations: Current State and Future Opportunities*, in *MIS Quarterly Executive*, 19, 4, 4, 2020; M. B. SCHRETTENBRUNNER, *Artificial-Intelligence-Driven Management*, in *IEEE Engineering Management Review*, 48, 2, 2020.

[830] A cominciare dall'ordinamento italiano, di cui è nota l'estrema complessità e il livello di dettaglio delle norme in materia di ordinamento giudiziario, cfr. F. DAL CANTO, *Lezioni di ordinamento giudiziario*, Torino, 2018.

[831] C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi cit.*, p. 3386-3387; F. DONATI,*Intelligenza artificiale e giustizia*, p. 418-421.

[832] Le tabelle aggiornate sono consultabili al sito: http://www.milanosservatorio.it/. In dottrina, cfr. il noto P. CENDON, A. NEGRO, *Danno biologico e tabelle milanesi*, Milano, 2011 e il più recente R. PARDOLESI, *Le nuove tabelle milanesi e il fascino discreto della para-normatività*, in *Danno e responsabilità*, 26, 4, 2021, p. 423-432.

loro elaborazione. Nonostantel'ipotesi goda, come già detto, del tendenziale consenso degli esperti, non risulta che essa sia stata, per ora, messa in pratica.[833]

- **Strumenti di intelligenza artificiale utilizzati nella fase delle indagini del procedimento penale**: è il caso, ad esempio, di captatori informatici di ultima generazione o di tecnologie avanzate per l'identificazione di sospetti, come sistemi di riconoscimento facciale o vocale. La diffusione di queste tecnologie nei vari paesi non è uniforme, anche a causa delle ingenti risorse richieste per la loro implementazione. Esse portano con sé il tema, complesso e ricco di sfaccettature, di quali garanzie procedurali applicarvi nei sistemi democratici, e in particolare se i vincoli imposti dai codici di procedura all'uso dei tradizionali mezzi di ricerca della prova siano sufficienti a disciplinare strumenti così invasivi della sfera individuale. Lo scenario risulta estremamente variegato e lo sviluppo tecnologico ancora acerbo, ma in via di estrema approssimazione può affermarsi che gli ordinamenti anglosassoni si dimostrano più liberali, riguardo al loro utilizzo, di quelli di *civil law*[834]. L'argomento, in ogni caso, esula dagli scopi di questo lavoro, ed è stato oggetto di studi approfonditi, negli ultimi anni, di diversi studiosi del procedimento penale, ai quali si rinvia[835]. In questa sede, ci si limita a ricordare che l'utilizzo per finalità di indagine relative ai reati cui è applicabile la disciplina sul mandato d'arresto europeo è una delle basi giuridiche che, nella Proposta di

[833] Non mancano, in ogni caso, sperimentazioni e ipotesi di lavoro a riguardo, anche nell'ordinamento italiano. Alcune, peraltro, risultano sorprendentemente risalenti: un esempio è rappresentato dal MoCAM, un sistema informatico per il calcolo dell'assegno divorzile, elaborato attraverso l'analisi dei precedenti del Tribunale di Firenze, comunque non riconducibile alla famiglia dell'apprendimento automatico e, in generale, all'intelligenza artificiale. Sul punto cfr., in particolare, F. DONATI,*Intelligenza artificiale e giustizia cit.*, p. 418-421.

[834]Cfr. ad esempio S. QUATTROCOLO, *Intelligenza artificiale e giustizia: nella cornice della Carta Etica Europea, gli spunti per un'urgente discussione tra scienze penali e informatiche cit.*, p. 10 ss.; per una panoramica dei principali utilizzi di tecnologie intelligenti nei sistemi penali anglosassoni, cfr. inoltre N. SCURICH, *The case against categorical risk estimates*, in*Behavioral Sciences & the Law*, 36, 5, 2018.

[835]Cfr. ad esempio U. PAGALLO – S. QUATTROCOLO, *The impact of AI in Criminal Law, and its Twofold Procedures cit.*;S. QUATTROCOLO, *Artificial Intelligence, Computational Modelling and Criminal Proceedings. A Framework for A European Legal Discussion*, Berlino, 2020; *Intelligenza artificiale e giustizia: nella cornice della Carta Etica Europea, gli spunti per un'urgente discussione tra scienze penali e informatiche cit.*;*Forecasting the future while investigating the past. The use of computational models in pre-trial detention decisions*, in*Revista Brasileira de Direito Processual Penal*, 7, 3, 2021; *Equo processo penale e sfide della società algoritmica*, in*BioLaw Journal – Rivista di BioDiritto*, 1, 2019, p. 135-144; *Risk assessment: sentencing o non sentencing?*, in A.A.V.V., *Giurisdizione penale, intelligenza artificiale ed etica del giudizio*, Milano, 2021, e*Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs. rischi e paure della giustizia digitale "predittiva"*, in *Cassazione penale*, 59, 4, 2019, p. 1748 ss.; F. BASILE, *Diritto penale e intelligenza artificiale*, in *Giurisprudenza italiana*, 172, 2019, p. 67-74; *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto penale e uomo*,10, 2019, p. 1-33; *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, in F. BASILE, M. CATERINI, S. ROMANO (A CURA DI), *Il sistema penale ai confini delle hard sciences: Percorsi epistemologici tra neuroscienze e intelligenza artificiale*, Pisa, 2021, p. 11-26; A. ZAVRŠNIK, *Criminal justice, artificial intelligence systems, and human rights cit.*;M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale cit.*; D. POLIDORO, *Tecnologie informatiche e procedimento penale cit.*; W. S. ISAAC, *Hope, Hype e Fear, The Promise and Potential Pitfalls of Artificial Intelligence in Criminal Justice cit.*

Regolamento dell'Unione Europea in materia di intelligenza artificiale, legittimano l'utilizzo di sistemi di riconoscimento biometrico in tempo reale[836].

Accanto a questi strumenti, devono menzionarsi i sistemi predittivi usati dalle autorità di polizia di diversi paesi, Italia compresa, al fine di valutare le possibilità di commissione di determinati crimini nel tempo e nello spazio e concentrare l'attività di indagine sugli *hotspot* così individuati[837]. Tali applicazioni dell'intelligenza artificiale, basate sull'analisi dei dati, sono ampiamente discusse nella letteratura giuridica e sociologica, per il rischio di esiti discriminatori per determinate comunità o gruppi sociali che, vivendo in aree caratterizzate da un più alto tasso di criminalità, finirebbero per essere oggetto di maggiori controlli da parte delle forze di polizia, chiamate ad agire in quelle zone dalla tecnologia[838]. Inoltre, è stato sottolineato il rischio che esse finiscano per autovalidare i propri risultati con un *bias* di conferma: la valutazione predittiva del sistema orienta le indagini in un determinato luogo, o verso una determinata categoria di persone; un reato viene commesso e i colpevoli sono subito individuati dalla polizia; il dato è utilizzato per l'autoapprendimento del sistema, che assegna al luogo un tasso di criminalità ancora più alto. Nel frattempo, vengono commessi reati anche in altre parti della città, o da soggetti il cui profilo è considerato portatore di un minor livello di rischio, ma le forze dell'ordine, concentrando i loro sforzi altrove, non ne vengono nemmeno a conoscenza (il tasso di denuncia per i reati minori contro il patrimonio, spesso, è molto basso). Il dato relativo, in

[836] L'art. 5 della Proposta di Regolamento, infatti, dispone: «[è vietato] l'uso di sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico a fini di attività di contrasto, a meno che e nella misura in cui tale uso sia strettamente necessario per uno dei seguenti obiettivi: i) la ricerca mirata di potenziali vittime specifiche di reato, compresi i minori scomparsi; ii) la prevenzione di una minaccia specifica, sostanziale e imminente per la vita o l'incolumità fisica delle persone fisiche o di un attacco terroristico; iii) il rilevamento, la localizzazione, l'identificazione o l'azione penale nei confronti di un autore o un sospettato di un reato di cui all'articolo 2, paragrafo 2, della decisione quadro 2002/584/GAI del Consiglio, punibile nello Stato membro interessato con una pena o una misura di sicurezza privativa della libertà della durata massima di almeno tre anni, come stabilito dalla legge di tale Stato membro».

[837] Alcuni esempi particolarmente noti sono rappresentanti dal software statunitense *PredPol*(ora noto come *Geolitica*) o da *KeyCrime*, sistema di *predictive policing* usato dalla Questura di Milano; cfr. inoltre il progetto *E-Security – ICT for knowledge-based and predictive urban security*, http://www.esecurity.trento.it/ (20 agosto 2022). In letteratura v. J.M. CAPLAN, L.W. KENNEDY, *Risk Terrain Modeling: Crime Prediction and Risk Reduction*, Berkeley, 2016; L.W. KENNEDY, J.M. CAPLAN, E.L. PIZA, *Risk Clusters, Hotspots and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies*, in *Journal of Quantitative Criminology*, 1, 2010, p. 339 ss.

[838] Cfr. ad esempio A. MEIJER, M. WESSELS, *Predictive policing: review of benefits and drawbacks,* in *International Journal of Public Administration*, 42, 12, 2019, p. 1031 ss.; W. PERRY, B. MCINNIS, C. C. PRICE, S. C. SMITH, J. S. HOLLYWOOD, *Predictive policing. The role of crime forecasting in law enforcement operations*, Washington, 2013; S. QUATTROCOLO, *Equo processo penale e sfide della società algoritmica cit.* eS. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs. rischi e paure della giustizia digitale "predittiva" cit.*; K. ALIKHADEMI, E. DROBINA, D. PRIOLEAU, *A review of predictive policing from the perspective of fairness*, in *Artificial Intelligence and Law*, 30, p. 1–17, 2022, https://doi.org/10.1007/s10506-021-09286-4; B. PEREGO, *Predictive policing: trasparenza degli algoritmi, impatto sulla privacy e risvolti discriminatori*, in *BioLaw Journal - Rivista di BioDiritto*, 2, 2020, p. 447-465; A. G. FERGUSON, *Policing Predictive Policing*, in *Washington University Law Review*, 94, 5, 2016, p. 1109 ss.; P. J. BRANTINGHAM, *The Logic of Data Bias and its Impact on Place-Based Predictive Policing*, in *Ohio State Journal of Criminal Law*, 15, 2017, p. 473 ss; P. J. BRANTINGHAM, M. VALASIK, G. O. MOHLER, *Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial*, in *Statistics and Public Policy*, 5, 1, 2018.

tal caso, non entra nel sistema, e la zona continua a non essere segnalata come possibile hotspot[839].

- **Strumenti di giustizia predittiva usati da operatori privati**: si tratta di tecnologie basate su elaborazione del linguaggio naturale e analisi dei dati che, attraverso l'esame di precedenti giurisprudenziali, aiutano avvocati e parti a valutare le possibilità di successo di una controversia, prevedendone il risultato finale. Si tratta di strumenti che hanno un crescente successo commerciale e che sono già realtà consolidate in alcuni ordinamenti: si pensi, ad esempio, all'impiego quotidiano da parte degli operatori del diritto statunitense, di *Lex Machina*, sistema di analisi dei precedenti sviluppato dal colosso dei servizi legali *Lexis Nexis*[840], o dal successo di software come *Prédictice* in Francia[841] o *Luminance* nel Regno Unito[842]. Questa tipologia di sistemi è solo la prima tra le applicazioni comprese in questo elenco della c.d. *giustizia predittiva*, ovvero l'insieme di applicazioni dell'intelligenza artificiale che, attraverso l'analisi dei dati, l'elaborazione del linguaggio naturale e l'apprendimento automatico punta allo sviluppo di modelli predittivi delle decisioni giudiziali.È anche quella che gli studiosi di diritto accolgono, tendenzialmente, con maggior favore, visto il miglioramento globale dell'assistenza legale ai privati che può derivarne, a fronte di un rischio per i diritti delle parti del processo che appare limitato. Maggiori discussioni, ad esempio, suscita l'applicazione degli stessi strumenti per la predizione di decisioni di un singolo tribunale, o la profilazione di un singolo giudicante, al fine di riscontrare difformità di giudizio tra corti o magistrati[843]. Tale applicazione dell'intelligenza artificiale è sembrata poter aiutare l'individuazione di *bias* ideologici e culturali di singoli giudici: ha acquisito particolare risonanza uno studio condotto in Francia, che, attraverso l'analisi dei risultati di un sistema di giustizia predittiva, ha riscontrato notevoli differenze nel tasso di accoglimento delle richieste di asilo e protezione internazionale tra diversi tribunali[844]. Allo stesso tempo, però, questi strumenti sono stati fortemente criticati, e accusati di poter dare origine a fenomeni di

[839]Ad evidenziare questa possibilità – talvolta indicata anche con l'espressione inglese *self-fulfilling prophecy* – è la stessa EUROPEAN COMMISSION FO THE EFFICIENCY OF JUSTICE (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment cit.*, p. 49-50.
[840]Cfr. https://lexmachina.com/ (20 agosto 2022).
[841]Cfr. https://predictice.com/fr (20 agosto 2022).
[842] Cfr.https://www.luminance.com/ (20 agosto 2022).
[843]Si veda l'analisi sul punto di S. PENASA, *Intelligenza artificiale e giustizia cit.*, p. 305-306. La più volte menzionata EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence cit.* inquadra la profilazione dei giudici tra gli utilizzi dell'intelligenza artificiale «*to be considered following additional scientific studies*», p. 66 ss.
[844]M. BENESTY, *L'impartialité de certains juges mise à mal par l'intelligence artificielle*, in *Village de la justice*, 25 marzo 2016, https://bit.ly/3qSaYuo (20 agosto 2022). Cfr. anche Y. MENECEUR, *L'intelligence artificielle en procès. Plaidoyer pour une réglementation internationale et européenne*, Bruxelles, 2020, p. 96-97.

forum shopping o minare l'autorità delle decisioni giudiziarie, e di rilevare, comunque, unicamente divergenze statistiche, la cui spiegazione potrebbe non risiedere in pregiudizi del giudicante: frequenza e tipologia dei casi relativi alla protezione internazionale in un tribunale situato vicino a un grande centro d'accoglienza, ad esempio, sono di certo diverse da quelle di un tribunale molto distante dai confini del paese. La Francia, come si vedrà, è giunta addirittura a vietare *ex lege* l'utilizzo di algoritmi per la profilazione dei singoli magistrati[845].

- **Strumenti di giustizia predittiva a supporto della decisione giudiziale**: applicazioni dell'intelligenza artificiale simili a quelle appena viste possono essere utilizzate come ausilio del magistrato giudicante, suggerendogli la decisione più coerente col corpo dei precedenti analizzati o evidenziando quali, tra questi ultimi, risultino più significativi sulla base dell'analisi dei dati. Questi sistemi possono, ovviamente, trovare impiego a supporto di ogni genere di valutazione giudiziale, non solo la sentenza che chiude definitivamente il giudizio[846]. L'ipotesi di un loro utilizzo genera, come si approfondirà, particolari discussioni, in ragione del rischio di c.d. distorsione dell'automazione, dell'ostacolo che tali sistemi rappresenterebbero per l'evoluzione giurisprudenziale, essendo fisiologicamente ancorati a precedenti del passato, e per la loro opacità, poiché si basano, frequentemente, sull'apprendimento automatico[847]. Inoltre, come già visto nella seconda parte del lavoro, il loro impiego pone questioni delicate relativamente alla possibilità di esiti discriminatori, anche in ambiti di rilevanza cruciale come le decisioni sulla libertà.L'utilizzo di questi strumenti, in particolare nei sistemi anglosassoni, in ogni caso, è un dato di realtà: possono citarsi, come esempi, i software COMPASS, in uso negli Stati Uniti, HART, per quanto riguarda il Regno Unito, o i sistemi predittivi al centro del caso *Ewert v. Canada*[848].

[845] L'intervento, come si dirà, è avvenuto con l'art. 33 della *Loi n. 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice*. Cfr. *infra*, p. 266 ss.

[846]Sull'impatto di tali sistemi sulle diverse dimensioni della decisione giudiziale, in particolare in ambito penale, si rimanda, in primo luogo, ai già citati studi di S. QUATTROCOLO, *Forecasting the future while investigating the past. The use of computational models in pre-trial detention decisions*, in*Revista Brasileira de Direito Processual Penal*, 7, 3, 2021; *Equo processo penale e sfide della società algoritmica*, in*BioLaw Journal – Rivista di BioDiritto*, 1, 2019, p. 135-144; *Risk assessment: sentencing o non sentencing?*, in A.A.V.V., *Giurisdizione penale, intelligenza artificiale ed etica del giudizio*, Milano, 2021.

[847] *Ex multis* cfr., da vari punti di vista, A. GARAPON – J. LASSÈGUE, *Justice digitale cit.*;U. PAGALLO – S. QUATTROCOLO, *The impact of AI in Criminal Law, and its Twofold Procedures cit.*;A. SANTOSUOSSO, *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritto cit.*; C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi cit.*; S. PENASA, *Intelligenza artificiale e giustizia: il delicato equilibrio tra affidabilità tecnologica e sostenibilità costituzionale cit.*; M. FASAN, *L'intelligenza artificiale nella dimensione giudiziaria cit.*;F. DONATI, *Intelligenza artificiale e giustizia cit.*; A. SIMONCINI, *L'algoritmo incostituzionale: l'intelligenza artificiale e il futuro delle libertà cit.*

[848]Si rimanda, in primo luogo, alla letteratura in commento ai noti casi *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016) con nota in *Harvard Law Review*, 130, 2017, p. 1530 ss. e *Ewert v. Canada*, 2018 SCC 30 [2018] 2 S.C.R. 165, citata

- **Strumenti di giustizia predittiva che sostituiscono la decisione giudiziale**: la sostituzione del giudice umano con sistemi avanzati di intelligenza artificiale è, generalmente, avversata da vari punti di vista e, nella letteratura giuridica di settore, non si rinvengono voci autorevoli che difendano la proposta[849]. Come si dirà, il rischio di possibili *bias*, errori e malfunzionamenti pare troppo alto, l'automazione totale sembra incompatibile con le garanzie processuali tipiche dei paesi democratici e la natura umana del giudice troppo importante anche per ragioni antropologiche, psicologiche e culturali. Il discorso è, invece, più complesso riguardo alle controversie di modesta entità e importanza, per le quali, spesso, il sistema giudiziario tradizionale non rappresenta uno strumento di tutela soddisfacente, per i costi d'accesso troppo alti e i tempi di celebrazione dei procedimenti troppo lunghi, anche quando siano previste, per tali cause, procedure particolarmente snelle e informali[850]. La proposta di automatizzare in tutto o in parte la loro decisione – magari solo per il primo grado di giudizio, o con procedure stragiudiziali che non precludano l'accesso al sistema formale –è spesso descritta come una delle possibli strategie per affrontare il problema, migliorando l'effettività della tutela garantita ai cittadini per vicende della vita quotidiana per le quali, allo stato dell'arte, spesso si rinuncia ad agire. L'idea non fa che coniugare le opportunità offerte dall'intelligenza artificiale e dal settore delle *Alternative Dispute Resolution* (ADR), spesso indicato come una delle possibili soluzioni alle inefficienze del sistema giustizia [851] . D'altronde, esistono già diverse procedure di c.d. *Online Dispute Resolution*(ODR), spesso predisposte da operatori privati, e,nonostante generalmente esse

---

nella seconda parte del lavoro. Cfr. in particolare L. HAN-WEI, L. CHING-FU, C. YU-JIE, *Beyond State v Loomis: Artificial Intelligence, Government Algorithmization and Accountability*, in *International Journal of Law and Information Technology*, 27, 2, p. 122–141; A.L. WASHINGTON, *How to Argue with an Algorithm: Lessons from the COMPAS-ProPublica Debate*, in *Colorado Technology Law Journal*, 17, 2018, p. 131 ss.; M. E. OLVER, *Some considerations on the use of actuarial and related forensic measures with diverse correctional populations*, in *Journal of Threat Assessment and Management,* 3, 2, 2016, p. 107–121; E. HILL, J. WOLFE, *Ewert v. Canada: Shining Light on Corrections and Indigenous People*, in *The Supreme Court Law Review: OsgoodÈs Annual Constitutional Cases Conference*, 94, 15, 2020, p. 391-413.

[849]Talvolta si assimilano a questa prospettiva alcune dichiarazioni di Daniel Kahneman, vincitore del *Nobel Memorial Prize* in *economic science* nel 2002, sull'auspicabilità di un utilizzo crescente dell'intelligenza artificiale nei procedimenti decisionali, poiché essa, al contrario degli esseri umani, non sarebbe soggetta a stanchezza e pregiudizi, cfr. J.N. MATHIAS, *Bias and Noise: Daniel Kahneman on Errors in Decision-Making*, Medium, 17 ottobre 2017, https://bit.ly/3BU6jhP (20 agosto 2022); C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi cit.*, p. 3374-3375.

[850]Cfr., da prospettive diverse e riguardo a ordinamenti distinti, A. J. SCHMITZ, *Expanding access to remedies through e-court initiatives*, in *Buffalo Law Review*, 67, 1, 2019, p. 89-165; A. C. BUDZINSKI, *Reforming service of process: an access-to-justice framework*, in *University of Colorado Law Review*, 90, 1, 2019; R. FABIO, *L'insostenibile complessità del processo: quale giustizia per gli small claims*, in P.G. MONATERI, A. SOMMA, *Patrimonio, persona e nuove tecniche di "governo del diritto". Incentivi, premi, sanzioni*, Napoli, 2009, p. 677-704; V. SANTO, *La società moderna e le nuove esigenze di accesso alla giustizia*, in *Studia Prawnoustrojove*, 24, 2014, p. 269-279.

[851]F. STEFFEK, H. UNBERATH, H. GENN, R. GREGER, C. MENKEL-MEADOW, *Regulating Dispute Resolution: ADR and Access to Justice at the Crossroads*, Oxford, 2013; J.T. BARRETT, J. BARRETT, *A history of alternative dispute resolution*, San Francisco, 2004.

digitalizzino la fase dell'accesso alla procedura, la cui gestione prevede, in seguito, l'intervento di esseri umani, non vi sono particolari ostacoli per l'implementazione di sistemi di intelligenza artificiale al loro interno che ne automatizzino il funzionamento[852]. L'Unione Europea ha dimostrato di credere particolarmente in tali strumenti, tanto da aver predisposto, ad esempio, un apposito portale informativosulle possibili strategie di ODR in materia di protezione del consumatore[853]. Allo stato dell'arte, in ogni caso, non constano, da parte delle autorità giudiziarie dei paesi membri, progetti di automazionedella risoluzione delle controversie che si basino, in tutto o in parte, sull'intelligenza artificiale: gli esempi principali già esistenti, come detto, prevedono sempre l'intervento di un essere umano, nel ruolo di arbitro o di mediatore/facilitatore[854]. Notizie di stampa, risalenti al 2019, relative al proposito della Repubblica dell'Estonia di digitalizzare e automatizzare il primo grado di giudizio, non sono state seguite dalla messa in pratica del progetto[855].

2.2. L'inquadramento giuridico dell'intelligenza artificiale nel settore della giustizia e la disciplina specifica emanata in Francia

Le applicazioni dell'intelligenza artificiale nel settore della giustizia sono soggette, prima di tutto, alle norme che disciplinano, in generale, le tecnologie intelligenti, in primo luogo in materia di trattamento dei dati personali. Sullo scenario europeo, l'art. 22 par. 1 del GDPR, già estensivamente commentato, sancisce il diritto a non essere sottoposti a una decisione giudiziaria totalmente automatizzata, posto che queste di certo producono, come richiesto dalla norma, «effetti giuridici» sull'individuo[856]. Il paragrafo successivo introduce, com'è noto, notevoli eccezioni, ad esempio il

[852] G. KAUFMANN-KOHLER, T. SCHULTZ, *Online Dispute Resolution: Challenges for Contemporary Justice*, Alphen, 2004; F.F. WANG, *Online Dispute Resolution: Technology, management and legal practice from an international perspective*, Oxford, 2008; D. CARNEIRO; P. NOVAIS; F. ANDRADE; J. ZELEZNIKOW; J. NEVES, *Online dispute resolution: an artificial intelligence perspective*, in *Artificial Intelligence Review*, 41, 2, 2014; C. RULE, *Online Dispute Resolution and the Future of Justice*, in *Annual Review of Law and Social Science*, 16, 1, 2020.

[853] Il portale è consultabile al link: https://bit.ly/2KMoGYL (20 agosto 2022). Per quanto riguarda il Regno Unito, cfr. UK CIVIL JUSTICE COUNCIL, *Online Dispute Resolution for low value civil claims*, 2015, https://bit.ly/3SgoxiJ (20 agosto 2022).

[854] Per una panoramica delle applicazioni principali in vari paesi del mondo cfr. ancora UK CIVIL JUSTICE COUNCIL, *Online Dispute Resolution for low value civil claims cit.*, p. 11-19, che passa in rassegna anche le principali applicazioni di sistemi simili nel settore privato (tipicamente da parte di società multinazionali per la gestione di reclami dei clienti).

[855] Il clamore mediatico attorno alla notizia è sorto da un articolo pubblicato nel 2019 sulla rivista online *Wired*, successivamente rilanciato dalla stampa di vari paesi, Italia compresa, cfr.E. NIILER, *Can AI Be a Fair Judge in Court? Estonia Thinks So*, Wired, 25 marzo 2019, https://bit.ly/2Tw9EdF (20 agosto 2022); A. DINI, *Debutta in Estonia il giudice-robot: le sentenze dall'intelligenza artificiale*, Corriere delle comunicazioni, 16 maggio 2019, https://bit.ly/2JmG5KL (20 agosto 2022). Il progetto, però, non è mai stato messo in pratica, né risultavano, sul punto chiare dichiarazioni ufficiali, tanto che, nel febbraio 2022, è arrivato un chiarimento ufficiale da parte del ministero della giustizia del paese baltico, cfr. REPUBLIC OF ESTONIA - JUSTIITSMINISTEERIUM, *Estonia does not develop AI Judge*, 16 febbraio 2022, https://bit.ly/3C46QOt (20 agosto 2022).

[856] Per il testo completo dell'art. 22 GDPR cfr. *supra*, p. 163 n. 543 e p. 202 n. 659.

consenso esplicito dell'interessato: l'eventuale automazione di una procedura giudiziaria sarebbe, allora, legittimata dall'adesione volontaria delle parti, qualora l'ordinamento continuasse a garantire anche la possibilità di adire il giudice in carne ed ossa. Inoltre, eccezioni al divieto di decisione interamente automatizzata possono essere introdotte dal diritto dei singoli stati membri: i paesi europei, Italia compresa, potrebbero, allora, istituire con legge sistemi automatizzati di risoluzione delle controversie in conformità all'art. 22 del Regolamento, col solo vincolo, già visto, di predisporre «misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato». Come si dirà al paragrafo successivo, la compatibilità col GDPR pare, comunque, un problema minore: l'automazione dell'attività decisionale dei magistrati, togati e onorari – da tenere distinta dallo sviluppo di eventuali procedure avanzate di *online dispute resolution* – solleva questioni più generali, relative alla compatibilità con le garanzie processuali più basilari.

La peculiarità del sistema giustizia, e i possibili sconvolgimenti derivanti dall'impiego, al suo interno, di tecnologie di intelligenza artificiale, sono ben rappresentati dalla circostanza che si tratta di uno dei settori dell'intelligenza artificiale che è già stato oggetto, in alcuni ordinamenti, di una disciplina specifica. È il caso, in particolare, della Francia, che, come già visto nel corso del lavoro, si stadimostrando particolarmente attiva nella regolazione delle nuove tecnologie.

Come già analizzato, la Francia è tra i paesi membri dell'Unione Europea che ha usufruito della possibilità di introdurre ulteriori condizioni di legittimità delle decisioni totalmente automatizzate, prevista dall'art. 22 par. 2 lett. b) del GDPR, modificando l'art. 47 della *Loi informatique et liberté* con le norme con cui, nel 2018, ha adeguato l'ordinamento interno all'entrata in vigore del Regolamento europeo[857]. Oltre alle innovazioni commentate al capitolo precedente, tale novella ha introdotto il divieto assoluto di automatizzare una decisione giudiziaria che contempli «une appréciation sur le comportment d'une personne». Il primo comma dell'art. 47, infatti, dispone: «aucune decisione de justice impliquant une appréciation sur le comportement d'une personne ne peut avoit fondement un traitement automatisé de donées à caractère personnel destiné a évaluer certains aspects de la personnalité de cette personne» [858] .La norma sembra prendere in considerazione, in particolare, strumenti di valutazione della personalità impiegati in giudizi prognostico-predittivi, come quelli al centro dei casi COMPAS e *Evert v. Canada*, mentre parrebbe, in astratto, non rientrare nel divieto un ipotetico sistema basato solamente sull'analisi, con tecniche di intelligenza artificiale, dei dati ricavati da precedenti giudiziari, utilizzato per giudizi d'altro genere. La delicatezza del contesto e la genericità della formulazione scelta –in fondo pressoché

[857] Come già riportato, si tratta in particolare della Loi n. 2018-493 du 20 juin 2018, *relative à la protection des données personnelles*.

[858] Per il testo completo della norma cfr. *supra*, p. 207 n. 683. In letteratura cfr. nuovamente G. MAUGERI, *Automated decision-making in the EU Member States: the right to explanation and other "suitable safeguards" in the national legislations cit.*, p. 13 ss.

ogni controversia tra persone fisiche contempla, almeno indirettamente, la valutazione di «certains aspects de la personnalité» - rendono plausibile, però, anche un'interpretazione estensiva.In ogni caso, l'ordinamento francese non è, finora, stato interessato da ipotesi di innovazioni tecnologiche che paventassero la possibile applicazione della norma.

Oltre all'appena commentata modifica della *Loi informatique et liberté*, la Francia ha introdotto, nel 2019, un'ulteriore interessante previsione, anch'essa di carattere restrittivo e riguardante l'intelligenza artificiale applicata al settore della giustizia. La vicenda, accennata al paragrafo precedente, ha origine con lo sviluppo, da parte dell'avvocato ed esperto di analisi dei dati parigino Michel Benesty, di un sistema che, elaborando i dati riguardanti le decisioni in materia di asilo e protezione internazionale, prevede il possibile esito di tal genere di cause in ogni tribunale, evidenziando notevoli differenze statistiche, nella percentuale media di accoglimento delle relative richieste, tra corti e singoli giudici. Tali discrepanze sono rapidamente interpretate come la dimostrazione di quanto *bias* ideologici e culturali influenzino il lavoro dei magistrati, anche in ambiti così delicati, e la questione attira l'attenzione della stampa, anche internazionale[859]. Viene creato addirittura un database gratuito e liberamente consultabile *online*, *SupraLegem*, con cui osservare il comportamento delle corti e valutare, di conseguenza, le possibilità di successo di una determinata istanza[860]. Il Legislatore è intervenuto con la *Loi n. 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice*, il cui art. 33 ha introdotto il divieto, presidiato penalmente, di riutilizzare i dati relativi all'identità di magistrati e ausiliari presenti nei provvedimenti giudiziali resi pubblici per «évaluer, analyser, comparer ou prédire leurs pratiques professionnelles réelles ou supposées»[861]. Il Conseil constitutionnel, adito in via diretta con *saisine parlamentaire*,ha sancito la conformità alla Costituzione francese di tali norme[862], considerandole una limitazione ragionevole al principio di pubblicità del processo (fatto derivare dagli artt. 6 e 16 della Dichiarazione dei diritti dell'uomo e del cittadino del 1789)[863]. Esse, infatti, contribuiscono in

[859]M. BENESTY, *L'impartialité de certains juges mise à mal par l'intelligence artificielle cit.* Si rimanda anche aY. MENECEUR, *L'intelligence artificielle en procès. Plaidoyer pour une réglementation internationale et européenne cit.*, p. 96-97.

[860]Una presentazione del servizio, enfatizzante le possibilità dischiuse dall'intelligenza artificiale, è ancora reperibile online: https://www.data.gouv.fr/fr/reuses/supra-legem/ (20 agosto 2022).

[861] L'art. 33 della Loi 2019-222 è l'unico articolo della *Section 3*, rubricata *Concilier la publicité des décisions de justice et le droit au respect de la vieprivée*. La norma ha novellato il *Code de justice administrative*(L. 10) e il *Code de l'organisation judiciaire*(L. 111-13), inserendovi due commi con la stessa formulazione: «Les données d'identité des magistrats et des membres du greffe ne peuvent faire l'objet d'une réutilisation ayant pour objet ou pour effet d'évaluer, d'analyser, de comparer ou de prédire leurs pratiques professionnelles réelles ou supposées. La violation de cette interdiction est punie des peines prévues aux articles 226-18,226-24 et 226-31 du code pénal, sans préjudice des mesures et sanctions prévues par la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés».

[862]Conseil Constitutionnel, Décision n° 2019-778 DC, 21 marzo 2019. Per dei commenti si rinvia al *Commentaire* a cura dell'ufficio studi del Conseil Constitutionnel, pubblicato unitamente alla sentenza, e nuovamente a S. PENASA, *Intelligenza artificiale e giustizia cit.*, p. 305-306.

[863]L'art. 6 della *Declaration* del 1789 recita: «La loi est l'expression de la volonté générale. Tous les citoyens ont droit de concourir personnellement ou par leurs représentants à sa formation. Elle doit être la même pour tous, soit qu'elle

modo proporzionato e ragionevole alla realizzazione del concorrente interesse a evitare fenomeni di *forum shopping* e pressioni sui magistrati, effetti indesiderati cui un sistema di profilazione dei giudici potrebbe facilmente portare[864]. L'innovazione legislativa e la sentenza costituzionale a conferma hanno, in ogni caso, suscitato reazioni discordanti: chi le ha valutate positivamente ha di fatto aderito all'impostazione del Conseil constitutionnel; chi le ha criticate, invece, sottolinea come la profilazione di giudici e tribunali con strumenti di intelligenza artificiale potrebbe rappresentare un prezioso strumento per incentivare la certezza del diritto, attraverso una maggiore coerenza tra la corti[865]. Deve evidenziarsi, inoltre, come le informazioni ricavate in tal modo non farebbero che contribuire all'aumento della trasparenza del sistema giustizia, ovvero uno dei valori che sono più di frequente menzionati tra i maggiori ostacoli al coinvolgimento di strumenti intelligenti nelle decisioni giudiziali. È singolare notare come tale interesse sia stato, invece, messo di fatto in secondo piano dal Legislatore francese al momento di valutare l'opportunità dell'impiego di strumenti d'intelligenza artificiale d'altro genere.

2.3. Le applicazioni dell'intelligenza artificiale nel settore della giustizia che non riguardano la decisione giudiziale: brevi cenni sull'impatto sui diritti fondamentali di alcune di esse

Le applicazioni dell'intelligenza artificiale nel settore della giustizia, al pari di quelle impiegate dalla pubblica amministrazione, sollevano questioni spinose per la protezione dei diritti fondamentali. È opportuno ribadire una considerazione preliminare: anche in ambito giudiziario, i problemi connessi all'implementazione di tecnologie intelligenti non riguardano solamente le applicazioni volte a replicare, assistere o sostituire l'attività decisionale del giudice, nonostante la letteratura giuridica abbia concentrato l'attenzione soprattutto su queste ultime. A esempio, i sistemi basati sull'elaborazione del linguaggio naturale con cui le grandi *law firm*, sempre più spesso, redigono e revisionano documenti legali pongono spinosi interrogativi riguardo alla necessità che essi siano "allenati" con dati di estrema qualità, al fine di limitare la possibilità, sempre presente, di

protège, soit qu'elle punisse. Tous les citoyens, étant égaux à ses yeux, sont également admissibles à toutes dignités, places et emplois publics, selon leur capacité et sans autre distinction que celle de leurs vertus et de leurs talents». Il notissimo testo dell'art. 16, invece, è: «Toute société dans laquelle la garantie des droits n'est pas assurée ni la séparation des pouvoirs déterminée, n'a point de Constitution».

[864]A sancirlo sono, in particolare, è i paragrafi 93 e 94 della sentenza: «93.En prévoyant que les données d'identité des magistrats et des membres du greffe figurant dans les décisions de justice mises à disposition du public par voie électronique ne peuvent faire l'objet d'une réutilisation ayant pour objet ou pour effet d'évaluer, d'analyser, de comparer ou de prédire leurs pratiques professionnelles réelles ou supposées, le législateur a entendu éviter qu'une telle réutilisation permette, par des traitements de données à caractère personnel, de réaliser un profilage des professionnels de justice à partir des décisions rendues, pouvant conduire à des pressions ou des stratégies de choix de juridiction de nature à altérer le fonctionnement de la justice. 94.Ces dispositions n'instaurent ainsi aucune distinction injustifiée entre les justiciables et ne portent pas d'atteinte contraire au droit à une procédure juste et équitable garantissant l'équilibre des droits des parties. Les griefs tirés de la méconnaissance, par ces dispositions, des articles 6 et 16 de la Déclaration de 1789 doivent donc être écartés».

[865]Cfr. ad esempio I. CONNETT, *France resists judicial AI revolution. France bans predicative analysis of caselaw. Does this protect or impede universal justice?*, in *Above the law*, 10 giugno 2019, https://bit.ly/3SlDCzw (20 agosto 2022).

errori. Anche il tema della concreta realizzabilità di un controllo da parte di professionisti umani sul loro funzionamento meriterebbe, probabilmente, una discussione più approfondita di quella che ha finora incontrato: la principale ragione dello sviluppo di tali sistemi, infatti, rimane proprio la loro capacità di esaminare un volume maggiore di documenti rispetto all'uomo, e più velocemente. Inoltre, l'utilizzo di tecnologie sempre più complesse per lo sviluppo di efficaci strategie di difesa in ambito giudiziale, come eventuali motori di ricerca intelligenti, dovrebbe essere analizzato anche dal punto di vista delle barriere d'accesso economiche a tali strumenti. Solamente i professionisti che potranno permettersi l'investimento, infatti, potranno usufruirne, con l'intuibile effetto di peggiorare ulteriormente la sproporzione che in molti casi esiste, già oggi, tra le difese di parti in giudizio in condizioni di forte squilibrio economico, rischiando di ridurre a una formula vuota il principio della parità delle parti garantito dalla difesa tecnica[866].

Tuttavia, le peculiarità della decisione giudiziaria, già evidenziate nel primo paragrafo di questo capitolo, non possono che portare a concentrare l'attenzione su di essa. Come già detto, l'esistenza dell'intero sistema giustizia, in fondo, è finalizzata alla formazione di tale decisione, e anche l'attività di consulenza stragiudiziale svolta dai grandi studi legali, analizzata poche righe sopra, trova nella volontà di scongiurare il rischio di finire di fronte a un giudice una delle ragioni principali della sua esistenza. Un'analisi delle conseguenze dell'implementazione crescente di tecnologie intelligenti nel sistema giudiziario dei paesi democratici, allora, deve riguardare, prima di tutto, le applicazioni che, in vario modo, possano interferire con le valutazioni dei magistrati. I paragrafi successivi affronteranno il tema nella prospettiva adottata in questo lavoro, basata sui diritti fondamentali, dei quali, d'altronde, il processo rappresenta la principale garanzia di effettività.

2.4. L'ipotesi del giudice algoritmico. L'incompatibilità coi diritti fondamentali e le ragioni metagiuridiche che rendono inaccettabile la sostituzione integrale del giudice umano

È bene trattare, in primo luogo, l'ipotesi dell'integrale sostituzione del giudice umano con strumenti basati sull'apprendimento automatico e l'elaborazione del linguaggio naturale. Si tratta di una prospettiva che, allo stato dell'arte, praticamente non trova sostenitori nella letteratura giuridica, tecnica e filosofica, né lo sviluppo di tecnologie di tal genere è mai stato assunto come obiettivo, nemmeno ipotetico a lungo termine, da un paese democratico[867]. Le ragioni dell'avversione alla

[866] Sul tema cfr. ad esempio S. CASERTA, *The sociology of the legal profession in the digital age*, in *International Journal of the Legal Profession*, 2021, https:doi.org/10.1080/09695958.2021.1920417 (21 agosto 2022).

[867] Riguardo alle notizie di stampa diffuse nel 2019, relative alle intenzioni della Repubblica dell'Estonia di sviluppare uno strumento algoritmico di risoluzione delle controversie – limitato, in ogni caso, a quelle di modesta entità – e alla successiva smentita iniziale, cfr. ancora E. NIILER, *Can AI be a fair judge in court? Estonia thinks so cit.* e REPUBLIC OF ESTONIA - JUSTIITSMINISTEERIUM, *Estonia does not develop AI Judge cit.* Come detto, la tesi della sostituzione del

totale automazione hanno a che fare coi diritti fondamentali, e in primo luogo con le garanzie processuali. Una decisione totalmente automatizzata, infatti, risulterebbe difficilmente motivabile e impugnabile, posta anche l'opacità che caratterizza, allo stato dell'arte, molti dei sistemi di apprendimento automatico più efficaci[868]. Eventuali giustificazioni dell'*output* ricavate con tecniche di *explainable artificial intelligence* sarebbero ben distanti, allo stato dell'arte, dalla complessità che comunemente caratterizza la motivazione dei provvedimenti giudiziali. Inoltre, pare doversi mettere in dubbio, più in generale, che le strategie di *explainable artificial intelligence* più diffuse possano portare a una spiegazione comparabile al ragionamento giudiziale. Parte delle tecniche più efficaci, infatti, come visto puntano a rendere più comprensibile l'output del sistema evidenziando quali variabili abbiano maggiormente contribuito alla sua elaborazione, evidenziando i nodi della rete più attivi, fornendo spiegazioni controfattuali, o con strategie di *input perturbation*. Si tratta di metodi che fanno luce sul funzionamento statistico del sistema, e puntano a fornire all'operatore umano la base di partenza per costruire da sé una giustificazione razionale dell'input che abbia la forma di un'argomentazione logica. Anche le tecniche che mirano a fornire spiegazioni basate su regole *if-then* - al netto dei loro limiti - difficilmente potranno raggiungere, coi loro risultati, una motivazione dell'output del sistema che non necessiti dell'intervento di un esperto umano per raggiungere risultati di ricchezza e complessità comparabile al ragionamento giuridico[869]. L'automazione della decisione giudiziale con l'impiego di reti neurali e altre tecniche di apprendimento automatico – allo stato dell'arte le più efficaci per l'analisi di moli di dati come un database di precedenti giudiziari – pare possibile, allora, solo al prezzo di sacrificare alcuni dei diritti di difesa più basilari.
Ulteriori argomenti a favore del mantenimento, almeno in parte, in mani umane della decisione giudiziale sono stati ricavati dall'ineliminabile possibilità di output erronei o discriminatori, mitigabile con la predisposizione di *dataset*d'allenamento di qualità, ma insita nella natura statistica delle tecnologie in esame[870]. I casi *Loomis v. Wisconsin* ed *Evert v. Canada*, già ampiamente

giudice umano con sistemi intelligenti è spesso attribuita a Daniel Kahneman, in virtù di alcune dichiarazioni sull'opportunità, in generale, di automatizzare i meccanismi decisionali umani, cfr. J.N. MATHIAS, *Bias and noise: Daniel Kahneman on errors in decision-making cit.*

[868] Cfr., in particolare, le considerazioni di V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *Discrimen*, 15 febbraio 2020, p. 14-17; C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi cit.*, p. 3380-3381 e F. DONATI, *Intelligenza artificiale e giustizia cit.*, p. 428-430. Si vedano anche, in senso più ampio, le approfondite considerazioni sul tema della motivazione delle decisioni pubbliche di A. SANTOSUOSSO, *Intelligenza artificiale e diritto cit.*, p. 100 ss.

[869] Fanno proprie queste considerazioni anche le EUROPEAN COMMISSION FO THE EFFICIENCY OF JUSTICE (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment*, p. 35 ss.

[870] Cfr. ad esempio M. LUCIANI, *La decisione giudiziaria robotica cit.*, p. 884-886; T. SOURDIN, *Judge v. Robot? Artificial Intelligence and Judicial decision-making cit.*, p. 1126 ss.; N. GESLEVICH PACKIN, Y. LEV-ARETZ, *Learning algorithms and discrimination*, in W. BARFIELD, U. PAGALLO (A CURA DI),*Research Handbook on the Law of Artificial Intelligence*, Cheltenham-Northampton (MA), 2018, p. 109 ss. Sulla natura statistica – e dunque, anche se in possibilità infinitesimali, sempre fallibile – dei sistemi di apprendimento automatico si rimanda ancora a S. QUINTARELLI, F. COREA, F. FOSSA, A. LOREGGIA, S. SAPIENZA, *AI: Profili etici. Una prospettiva etica sull'intelligenza artificiale cit.*, 195 ss.

analizzati, hanno messo in luce i pericoli per l'effettività del principio di eguaglianza che possono derivare dall'utilizzo di determinati strumenti di intelligenza artificiale a supporto dell'attività decisionale dei magistrati. È evidente che tali rischi non farebbero che peggiorare qualora tale attività fosse totalmente automatizzata, rendendo, peraltro, molto meno agevole identificare in tempi brevi le disfunzioni del sistema che li causino, poiché questo sarebbe sottratto al fondamentale controllo di un operatore umano esperto.

Alcuni, tra gli studiosi di diritto che si sono occupati più a fondo del tema, hanno acutamente osservato che anche la decisione giudiziale si comporta come una *black-box* e ha, talvolta, esiti discriminatori, pur senza trarne la conclusione che l'automazione totale sia auspicabile[871]. Invero, come già analizzato, sono ormai numerosi gli studi che mostrano come la motivazione razionale di una decisione umana sia, in realtà, in molti casielaborata *ex-post*, e che il momento della scelta avvenga sulla base di pulsioni emotive e razionali delle quali, spesso, la persona neppure dimostra coscienza[872]. L'esistenza di una netta cesura tra elaborazione della decisione e formulazione delle sue giustificazioni è comunemente accettata anche dalla migliore dottrina giuridica[873]. I sistemi processuali che non prevedono la possibilità di opinione dissenziente per i membri di un contesto giudicante, inoltre, rendono di fatto inaccessibile parte del procedimento motivazionale, presentando come unanime una valutazione cui invece si è giunti solo al termine di un confronto dialogico molto acceso[874]. È evidente, inoltre, che i provvedimenti dei magistrati umani possono avere esiti discriminatori, se non altro perché, in molti casi, è attraverso l'analisi dei loro precedenti che le tecnologie intelligenti giungono al medesimo risultato[875].

Queste considerazioni, per quanto corrette, non portano a mutare la posizione di rigettodella delega totale all'algoritmo della decisione giudiziale: in primo luogo perché, come già analizzato trattando del diritto alla spiegazione, essere umano e macchina rimangono ontologicamente distinti. L'insondabilità degli stati mentali altrui e delle reali ragioni delle loro azioni non è un problema

[871] Cfr. in particolare A. SANTOSUOSSO, *Intelligenza Artificiale e Diritto cit.*, p. 100 ss., su cui C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi cit.*, p. 3380 ss., che distinguono due fasi nella valutazione giudiziale: il *context of discovery* (la decisione vera e propria) e il *context of justification* (la motivazione a supporto). Significativamente, ambo gli autori richiamanoP. CALAMANDREI, *Processo e democrazia*, Padova, 1954, p. 101, che già si riferiva alla motivazione come all'«apologia che il giudice elabora a posteriori della decisione stessa». Si vedano anche i rilievi di S. ARDUINI, *La "scatola nera" della decisione giudiziaria: tra giudizio umano e giudizio algoritmico*, in *BioLaw Journal – Rivista di BioDiritto*, 2, 2021, p. 463 ss. In generale, su come le spiegazioni fornite dagli esseri umani alle loro azioni e convinzioni siano spesso giustificazioni postume di procedimenti istintuali e irrazionali, e sul paragone tra essere umano e sistemi artificiali c.d. *black-box*, si rimanda a quanto ampiamente esposto *supra*, p. 185 ss..

[872] Cfr. ancora *supra*, p. 185 ss.

[873] Si veda nuovamente A. SANTOSUOSSO, *Intelligenza artificiale e diritto cit.* e, in generale,C. P. GUTHRIE; J. J. RACHLINSKI; A. J. WISTRICH, *Inside the Judicial Mind*, in *Cornell Law Faculty Publications,* paper 814*,* 2001 http://www.ssrn.com/abstract=257634 (21 agosto 2022).

[874] Cfr. ancora A. SANTOSUOSSO, *Intelligenza artificiale, conoscenze neuroscientifiche e decisioni giuridiche*, in *Teoria e critica della regolazione sociale*, 1 2021, p. 189 ss.

[875] Tra i molti possibili esempi di decisioni di giudici umani influenzate da pregiudizi, si richiamano i risultati, già citati e significativamente ottenuti attraverso l'analisi dei dati con strumenti di intelligenza artificiale, di M. BENESTY, *L'impartialité de certains juges mise à mal par l'intelligence artificielle cit.*

perché assumiamo che essi somiglino ai nostri, in base al valido argomento di riconoscerci come membri della stessa specie. Allo stesso modo, ci sembra fisiologica l'imperfezione dei loro giudizi, perché è anche la nostra. Mantenere questa impostazione nei confronti di un'applicazione tecnologica è molto più complicato, perché di essa percepiamo immediatamente l'alterità, ci risulta meno comprensibile e perché associamo all'automazione livelli di rapidità, precisione ed efficienza maggiori di quelli ottenibili dall'essere umano[876]. Queste osservazioni valgono, in generale, per ogni forma di delega alla tecnologia, e sono valide a maggior ragione per la decisione giudiziale, riguardo alla quale, peraltro, sono supportate da elementi culturali, anche extragiuridici, molto risalenti e di estrema rilevanza.

L'idea che giudicante e giudicato debbano condividere lo stesso terreno, infatti, è ben radicata nella tradizione giuridica occidentale. Essa è presente già nella *Magna Charta Libertatum* del 1215, che, nel porre le basi di una delle garanzie giuridiche più basilari dello stato democratico odierno, sanciva per la prima volta che un uomo potesse essere processato e imprigionato solamente in base al «lawful judgment of his equals»[877]. In modo simile, i provvedimenti giudiziari, in Italia e in molti altri paesi, sono emanati dai giudici "in nome del popolo", ovvero, in ultima analisi, dell'intera comunità dei loro simili. Pare difficile pensare che una macchina possa veramente considerarsi un pari dell'essere umano che fosse posta a giudicare, o pronunciare una sentenza in nome di un popolo del quale di certo non fa parte. Così come pare dubbio, limitandoci a un solo ulteriore esempio, tratto dalla Costituzione italiana, che un algoritmo possa considerarsi il «giudice naturale» di qualcuno[878]. Se, dunque, l'idea che giudicante e giudicato debbano appartenere alla medesima categoria è ben presente nella nostra cultura, prima che nella nostra tradizione giuridica, la conclusione che essi debbano prima di tutto riconoscersi reciprocamente come membri della specie umana, a pena del rovesciamento di una delle categorie concettuali con cui interpretiamo il mondo, pare obbligata[879].

---

[876] Sulla radicale differenza tra le sensazioni suscitate dall'interazione con un essere umano e con un sistema intelligente, si rimanda a B. ERB, *Artificial intelligence & the theory of mind cit.*; T. ARAUJO, *In AI we trust? Perceptions about automated decision-making by artificial intelligence cit.*

[877] L'espressione proviene dal punto 39 della Magna Charta, che, nella traduzione più accreditata in inglese moderno, recita: «No free man shall be seized or imprisoned, or stripped of his rights or possessions, or outlawed or exiled, or deprived of his standing in any way, nor will we proceed with force against him, or send others to do so, except by the lawful judgment of his equals or by the law of the land». Il paragone con l'ipotesi di un futuro giudice algoritmico è di C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi cit.*, p. 3384-3385 e *AI and constitutionalism: the challenges ahead cit.*

[878] Così ad esempio F. DONATI, *Intelligenza artificiale e giustizia cit.*, p. 429 e ancora C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi cit.*, p. 3381 ss.

[879] Argomentano l'esclusività umana di determinate caratteristiche necessarie al giudizio (come creatività, libero apprezzamento, sensibilità, capacità di riconoscere e proteggere i propri errori) che portano a considerare come "ontologicamente umana" la professione del giudice, da vari punti di vista, tra gli altri, A. D'ALOIA, *Il diritto verso "il mondo nuovo". Le sfide dell'Intelligenza Artificiale*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019; M. LUCIANI, *La decisione giudiziaria robotica cit.*; T. SOURDIN, R. CORNES, *Do Judges Need to Be Human? The Implications of*

Infine, preme sottolineare che, allo stato dell'arte, le uniche ipotesi di totale automazione della decisione di una controversia che sembrano avere la concreta possibilità di essere messe in atto riguardano le cause di modesta entità. È doveroso evidenziare che ognuna delle questioni appena viste, a cominciare dalla possibilità di *bias* ed errori e dalla difficoltà di ricavare dall'algoritmo una motivazione paragonabile all'argomentazione giudiziale, esiste, inalterata, anche riguardo a tali controversie. La possibilità di una delega completa alla macchina delle decisioni in tali materie, con l'effetto di sottrarre le relative cause a ogni forma di controllo del giudice umano, è, allora, da rigettare. Diversa, invece, pare l'opportunità di predisporre procedure di risoluzione alternative automatizzate, che non precludano l'accesso all'autorità giudiziaria tradizionale né lo rendano più oneroso o difficoltoso, e rappresentino, al contempo, uno strumento efficiente, innovativo ed efficace per ottenere giustizia in tempi brevi[880]. L'ipotesi, in buona sostanza, dovrebbe portare a un incremento del diritto all'effettività della tutela giudiziale e non alla preclusione del diritto di accesso alla giustizia per persone in condizioni di difficoltà economica o culturale, o non in grado di interfacciarsi con i necessari strumenti tecnologici.

2.5. Intelligenza artificiale a supporto della decisione giudiziale e vecchi e nuovi diritti fondamentali

L'ipotesi dell'impiego di strumenti di intelligenza artificiale a supporto del giudice umano è l'unica, come già visto, ad aver avuto attuazione pratica. Il suo potenziale impatto sui diritti fondamentali, in primo luogo sul principio di eguaglianza e non discriminazione, è, peraltro, già stato messo in luce da alcune vicende discusse dalle corti di vari paesi, come i già più volte menzionati casi *Loomis* ed *Evert*.

Dal punto di vista delle garanzie procedurali, il permanere di un magistrato umano, titolare della decisione, permette di superare le ovvie obiezioni, viste al paragrafo precedente, in materia di garanzie processuali: una motivazione del provvedimento è senza dubbio presente, poiché il giudice la redigerà come farebbe in assenza dell'ausilio tecnologico; tale provvedimento motivato risulterà

Technology for Responsive Judging, in T. SOURDIN, A. ZARISKI (A CURA DI), *The Responsive Judge: International Perspectives*, Singapore, 2018, p. 87-119.

[880]Un sistema giudiziario in cui l'intelligenza artificiale abbia un ruolo soprattutto nella risoluzione di controversie civili di minor entità, attraverso lo sviluppo di nuove opzioni per la risoluzione delle controversie, in primo luogo ODR, senza mai giungere alla completa sostituzione del giudice umano, è ipotizzato, tra gli altri, da T. SOURDIN, *Judge v. Robot? Artificial Intelligence and Judicial decision-making cit.*, p. 1131 ss. Si segnala che la più volte citata EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment cit.* include gli utilizzi dell'intelligenza artificiale con finalità di *Support for alternative dispute settlement measures in civil matters* e *Online dispute resolution* tra i *Possible uses, requiring considerable methodological precautions*, p. 64-65. In particolare, essa invita a garantire, in ogni caso, la possibilità di intervento di una «trained third party (mediator using not only techniques but maybe scales as calculated above, or a lawyer) would appear to be the most appropriate solution at this stage» (p. 64). Si confronti, inoltre, l'interessante opinione di F. DONATI, *Intelligenza artificiale e giustizia cit.*, che argomenta che l'intelligenza artificiale potrebbe rivelarsi un ausilio prezioso per il c.d. "filtro in appello" nel procedimento civile italiano.

impugnabile; le modalità di esercizio del diritto d'azione sembrano poter permanere inalterate, posto che non è necessario che la parte, pubblica o privata, si interfacci direttamente col sistema. Allo stesso modo, la presenza del giudice umano dovrebbe assicurare che il procedimento si svolga regolarmente, e i diritti e gli interessi coinvolti vi trovino una composizione giusta ed efficace. In astratto, le possibilità dischiuse dal sistema intelligente in termini di efficienza dovrebbero incrementare la capacità del processo di rappresentare effettivamente il luogo privilegiato della tutela dei diritti fondamentali che sempre dovrebbe essere.

Il discorso, però, si complica enormemente quando si esaminino le caratteristiche delle tecnologie impiegate a supporto del giudice, in primo luogo dal punto di vista dei nuovi diritti teorizzati in questo lavoro. È stato più volte ripetuto, infatti, che le applicazioni dell'intelligenza artificiale più comunemente utilizzate a tale scopo consistono in strumenti basati sull'analisi dei dati con sistemi di apprendimento automatico, in primo luogo le reti neurali[881]. La conseguenza che ne può derivare è stata, anch'essa, già analizzata: in molti casi queste tecnologie si comportano come *black-box*, rendendo difficile comprenderne gli stati interni e spiegare razionalmente i loro risultati finali. Ciò, ovviamente, può generare problemi non irrilevanti quando siano utilizzate a supporto dell'attività decisionale del giudice umano.Innanzitutto, vi è il rischio, già menzionato, definito *effet moutonnier* dal magistrato e studioso francese Antoine Garapon e "distorsione dell'automazione" dalla Proposta di Regolamento in materia di intelligenza artificiale presentata dall'Unione Europea nell'aprile del 2021[882]. L'interazione con una tecnologia che automatizza, in tutto o in parte, un'attività della quale si era in precedenza incaricati *in toto* spinge a non mettere in discussione il comportamento di quest'ultima, per la percezione di oggettività che, anche a livello inconscio, circonda la tecnologia e per l'erosione delle competenze necessarie allo svolgimento di tali compiti causata dall'inattività (*deskilling*)[883]. La circostanza che tale strumento avanzato consista in un sistema del quale si

[881] Si rimanda alla rassegna delle principali tecnologie di giustizia predittiva svolta *supra*, p. 262 ss.

[882] A. GARAPON, J. LASSEGUE, *Justice digitale*, p. 239. L'espressione «distorsione dell'automazione» compare all'art. 14 par. 4 lett. b) della Proposta di Regolamento, che così la definisce: «restare consapevole della possibile tendenza a fare automaticamente affidamento o a fare eccessivo affidamento sull'output prodotto da un sistema di IA ad alto rischio, in particolare per i sistemi di IA ad alto rischio utilizzati per fornire informazioni o raccomandazioni per le decisioni che devono essere prese da persone fisiche».

[883] Cfr. ad esempio i già citati J. LU, *Will Medical Technology Deskill Doctors?*, in *International Education Studies*, 9, 7, 2016, p. 130–134; J. LEVY, A. JOTKOWITZ, I. CHOWERS, *Deskilling in Ophthalmology Is the Inevitable Controllable?*, in *Eye*, 33, 3, 2019, p. 347–348; S. DE PAOLI, *Automatic-Play and Player Deskilling*, in *MMORPGs,Game Studies*, 13, 1, 2013, relativi a evidenze scientifiche di un possible deskilling sul lungo termine, a causa dell'automazione, di medici e giocatori di famosi videogiochi online. L'eventuale *deskilling* subito dai magistrate risulterebbe, peraltro, particolarmente problematico da individuare, perchè la qualità delle loro decisioni sembra più difficilmente misurabile rispetto alla percentuale di successi negli interventi di un chirurgo esparto, o di mosse corrette da parte di un giocatore. Sulle conseguenze di questo genere di *moral deskilling* cfr., da vari punti di vista, S. VALLOR, *The future of military virtue: autonomous systems and the moral deskilling of the military*, in *2013 5th International Conference on Cyber Conflict (CYCON 2013)*, 2013, https://bit.ly/3qU4URV (21 agosto 2022); S. VALLOR, *Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character*, in *Philosophy & Technology*, 28, 1, 2015, p. 107-124; K.L. MOSIER, L.J. SKITKA, *Human Decision Makers and Automated Decision Aids: Made for Each Other?*, in

conosce l'altissima efficacia, strutturalmente superiore a quella raggiungibile dall'essere umano, e di cui non sia possibile comprendere le modalità di funzionamento, né costruire un percorso logico-razionale a supporto degli output, ovviamente, incide sul fenomeno. Discostarsi dai risultati proposti da una tecnologia di tal genere pare enormemente complesso, e affidarsi acriticamente ad essa terribilmente semplice e immediato. Calando il discorso sull'ambito che si sta esaminando, in caso di disaccordo col sistema, il giudice umano come potrebbe superare il dubbio di essere in errore, posto che la macchina analizza e ricorda molti più precedenti? Sembra un compito proibitivo, in particolare, quando la questione abbia ad oggetto valutazioni dall'accentuato carattere prognostico, come quelle che, in ambito penale, implicano un giudizio sulla personalità. Tali decisioni, infatti, si basano, in modo accentuato su indicatori probabilistici e massime d'esperienza ricavatedall'analisi di casi passati, per la quale, però, il giudice può attingere solo alla sua esperienza, mentre il sistema intelligente a un database più ampio di diversi ordini di grandezza. Non è un caso, probabilmente, che proprio sistemi destinati al supporto di tali valutazioni siano stati tra i primi a suscitare discussioni[884]. Inoltre, il rischio di sostanziale asservimento del magistrato alla tecnologia che lo assiste rende solo apparente il rispetto delle garanzie processuali commentate in precedenza: la motivazione della decisione, in particolare, consisterà nella semplice giustificazione *ex-post* dell'indicazione dell'algoritmo, che non risulta spiegabile e che, proprio per questo, il giudice ha imparato a non mettere in discussione[885]. La possibilità di impugnazione rimarrebbe, invece, inalterata, ma è ragionevole presumere che ogni ipotesi d'appello tenterebbe di mettere in discussione, in primo luogo, proprio le modalità di utilizzo della tecnologia utilizzata a supporto, col risultato che essa finirebbe per dominare anche i successivi gradi di giudizio (qualora essa non fosse, semplicemente, utilizzata anche in tale sede).

Anche l'utilizzo dell'automazione come ausilio al giudice, e non al fine di sostituirlo integralmente, allora, può portare al fallimento del processo per le finalità a cui è preposto e al conseguente, significativo deterioramento di diritti e prerogative individuali. I nuovi diritti proposti in questo studio sembrano, allora, acquisire particolare importanza.Il coinvolgimento di tecnologie intelligenti dovrebbe, ovviamente, essere reso noto alle parti del procedimento, in modo che esse possano esercitare con pienezza, anche da quel punto di vista, il loro diritto di difesa. A venire in gioco sono,

---

Automation and Human Performance: Theory and Applications, Boca Raton, 1996, https://bit.ly/3Lw2wKC (21 agosto 2022).

[884] L'ovvio riferimento è alle vicende di discriminazione algoritmica già giunti di fronte alle corti di vari paesi analizzati nella seconda parte, e in particolare ai più volte menzionati casi *Loomis* ed *Evert*.

[885] Come già analizzato, *context of discovery* e *context of motivation* vanno tenuti distinti, e potrebbe affermarsi che anche la motivazione della decisione presa dal giudice umano, senza il supporto di alcun sistema tecnologico, consista in una giustificazione *ex post*. Ad essere decisiva, però, è la già rilevata alterità dell'essere umano rispetto all'elemento tecnologico, che anche in quest'ultimo caso risulta, di fatto, l'unico decidente. Cfr. ancora A. SANTOSUOSSO, *Intelligenza Artificiale e Diritto cit.*, p. 100 ss.; B. ERB, *Artificial intelligence & the theory of mind cit.*; T. ARAUJO, *In AI we trust? Perceptions about automated decision-making by artificial intelligence cit.* e in generale supra, p. 185 ss.

però, soprattutto il diritto alla spiegazione e il diritto al controllo umano, nella loro intrinseca connessione. Le specificità della decisione giudiziaria – che, come detto, pare opportuno mantenere sotto il pieno controllo dell'essere umano, anche per ragioni metagiuridiche – impongono un'applicazione estremamente rigorosa del requisito della spiegabilità, che pare difficilmente sacrificabile in nome di altri interessi, connessi a un'eventuale maggior efficacia del sistema[886]. Il ridimensionamento del ruolo del giudice umano che ne conseguirebbe, infatti, non pare mai accettabile. Sarà necessario, allora, esaminare rigorosamente le tecnologie avanzate che si ipotizzasse di utilizzare a supporto del giudice in attività considerabili, in senso ampio, decisionali o valutative, per verificare che esse, per le modalità di sviluppo impiegate o l'applicazione di tecniche di *explainable artificial intelligence* di particolare qualità ed efficacia, presentino un livello di spiegabilità sufficiente. Solo così, infatti, sarà possibile evitare le conseguenze spiacevoli e disumanizzanti a cui un livello di automazione troppo avanzato può condurre nel settore della giustizia e garantire che il processo si svolga nel pieno controllo del giudice umano a cui tale compito è demandato, un diritto fondamentale sia delle parti private che di quest'ultimo.

Da questo punto di vista, l'approccio scelto dalla Proposta di Regolamento dell'Unione Europea in materia di intelligenza artificiale potrebbe risultare adatto allo scopo, in ragione del particolare rigore di alcune delle norme tecniche dettate per le tecnologie ad alto rischio, in particolare quanto previsto all'art. 14 in materia di *human oversight*, già analizzato nel dettaglio nel corso del lavoro[887]. Tuttavia, non può non segnalarsi come la decisione giudiziaria presenti peculiarità evidenti, che la distinguono da ogni altro ambito, e che forse meriterebbero una disciplina specifica, in grado di definire il regime applicabile alle diverse applicazioni tecnologiche e situazioni di riferimento che possano venire in gioco. Sistemi come i *risk-assessment tool* in ambito penalistico, o l'utilizzo di software di giustizia predittiva per la profilazione di singoli magistrati, infatti, meritano, probabilmente, un inquadramento giuridico a sé stante, vista la diversità dei problemi cui possono portare. Da questo punto di vista, l'ordinamento francese, che ha scelto di disciplinare in modo particolarmente precoce alcune applicazioni specifiche dell'intelligenza artificiale nel sistema giustizia, sembra poter rappresentare un modello da imitare, al netto dell'impostazione, come visto notevolmente restrittiva, per ora adottata da esso, che, in parte, potrebbe anche non essere condivisa.

[886] Cfr. ad esempio C. RUDIN, *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead*, in *Nature Machine Intelligence*, 1, 5, 2019, p. 206-215.

[887] Si specifica che l'Annex III della Proposta di Regolamento include tra le tecnologie ad alto rischio, ad esempio, «i sistemi di IA destinati a essere utilizzati dalle autorità di contrasto per effettuare valutazioni individuali dei rischi delle persone fisiche al fine di determinare il rischio di reato o recidiva in relazione a una persona fisica o il rischio per vittime potenziali di reati» o i «sistemi di IA destinati ad assistere un'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti»

3. Intelligenza artificiale, medicina e diritti fondamentali

3.1. Cenni sulle principali applicazioni dell'intelligenza artificiale in ambito sanitario e alcuni possibili sviluppi futuri

Il settore della sanità, il terzo dei campi di applicazione dell'intelligenza artificiale che si è scelto di approfondire in questa ultima parte del lavoro, è da sempre uno dei terreni di elezione dell'innovazione tecnologica. La rilevanza primaria degli interessi coinvolti, infatti, fa sì che le possibilità di sviluppi nel settore trovino adeguati finanziamenti economici più facilmente che in altri ambiti[888]. Inoltre, le possibilità dischiuse dalla tecnologia nel corso del XX secolo, sia per l'indagine diagnostica (si pensi ai sistemi ECG o agli apparecchi radiologici)[889] che per l'attività di cura e assistenza al paziente (un esempio particolarmente noto, e relativamente risalente, è rappresentato dai macchinari per l'emodialisi)[890] hanno reso l'ambito sanitario uno dei settori in cui l'interazione tra professionista umano e macchine complesse è più frequente. Le tecnologie basate sull'intelligenza artificiale si sono facilmente inserite in questo generale quadro di apertura nei confronti della tecnologia. Ad esempio, due tra i primi sistemi esperti, MYCIN[891] e CASNET[892], erano destinati all'attività medica, rispettivamente all'individualizzazione del trattamento antibiotico e al monitoraggio del glaucoma. Intuendo i progressi raggiungibili con la loro elaborazione con strumenti avanzati, la digitalizzazione di pubblicazioni mediche ed evidenze cliniche è stata particolarmente precoce rispetto ad altri settori – un ovvio esempio, in molti paesi, è costituito dalle banche dati giurisprudenziali – tanto che l'inizio della compilazione, a cura della

[888] Preme chiarire che quanto affermato è vero solo con importanti distinguo: il finanziamento privato a iniziative di ricerca e sviluppo in ambito tecnologico-sanitario segue, ovviamente, in primo luogo logiche di mercato; l'intervento pubblico a sostegno della ricerca scientifica, anche in ambito biomedico, spesso e in vari paesi, è considerato largamente insufficiente dalla maggior parte dei commentatori specializzati; la difficoltà a reperire finanziamenti in settori dell'attività sanitaria che si rivelino fallimenti del mercato, nonstante gli interessi vitali coinvolti (si pensi alla cura di determinate malattie rare) è ben nota.

[889] La "scoperta" dell'elettrocardiogramma si fa generalmente risalire a A.D. WALLER, *A Demonstration on Man of Electromotive Changes accompanying the Heart's Beat*, in *The Journal of Physiology*, 8, 5, 1887, p. 229-234, mentre le prime concrete applicazioni cliniche al secondo decennio del '900, cfr M. ALGHATRIF, J. LINDSAY,*A brief review: history to understand fundamentals of electrocardiography*, in *Journal of Community Hospital Internal Medicine Perspectives*, vol. 2, n.° 1, 2012, p. 14383 ss. La moderna radiologia, invece, ha origine a cavallo tra XIX e XX secolo, negli studi condotti, tra gli altri, da Nicola Tesla, William Crookes, Hermann von Helmhotz, Philipp von Leonard, Heinrich Hertz e Wilhelm Röntgen, che per questo vinse il premio Nobel, cfr. A.M.K. THOMAS, *The history of radiology*, Oxford, 2013, p. 1-11.

[890] Com'è noto, la possibilità di utilizzare macchinari per l'emodialisi per il trattamento dell'insufficienza renale cronica si deve alle tecnologie introdotte nel 1960 all'ospedale di Seattle dal direttore della divisione di nefrologia dell'università di Washington Belding Hibbard Scribner. La limitata disponibilità di tali tecnologie portò, nel primo periodo, alla necessità di selezionare i pazienti cui garantire l'accesso al trattamento, una scelta al tempo delegata a un comitato composto da persone di varia estrazione sociale, che si guadagnò, in breve, l'appellativo di *Life or Death Committee*, cfr. S. ALEXANDER, *They decide who lives, who dies. Medical miracol and a moral burden*, in *Life*, 9, 1972, p. 102 ss; C. CASONATO, *Introduzione al biodiritto cit.*

[891] Cfr. SHORTLIFFE, B.G. BUCHANAN, *A model of inexact reasoning in medicine*, in *Mathematical Biosciences*, 23, 3-4, 1975, p. 351-379; B.G. BUCHANAN, E. H. SHORTLIFFE, *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project*, Reading (USA), 1994. Vedi, inoltre, *supra*, p. 40 ss.

[892]S. WEISS, C. KULIKOWSKI, A. SAFIR, *Glaucoma consultation by computer*, in *Computers in Biology and Medicine*, 8, 1, 1978, p. 25-40.

National Liberary of Medicine, del notissimo database *PubMed*, disponibile online dal 1996, risale ai primi anni '60[893]. Gli avanzamenti nella robotica hanno avuto, com'è noto, ampia applicazione in medicina, e in primo luogo nell'attività chirurgica, nella quale da tempo rivestono, in diversi ambiti, un ruolo insostituibile[894].

L'intelligenza artificiale, dunque, è impiegata in ambito sanitario sin dalla comparsa delle sue prime applicazioni commerciali e industriali. Tuttavia, in epoca più recente, l'avvento delle reti neurali, e in particolare del *deep learning*, ha rappresentato una rivoluzione per il settore eguagliata in pochi altri contesti[895]. Il campo, infatti, si è dimostrato particolarmente adatto allo sviluppo di tale famiglia di tecnologie, per varie ragioni: in primo luogo, la natura della scienza medica, caratterizzata da fortissima specializzazione e dal ruolo decisivo, per i suoi risultati, dell'esame di una mole elevata di casi clinici; in secondo luogo, l'ampia presenza di dati digitalizzati; infine, la già menzionata varietà e solidità delle attività di ricerca e sviluppo. Le applicazioni delle reti neurali apparse dirompenti hanno riguardato, in particolare, l'attività diagnostica: sistemi di riconoscimento e interpretazione delle immagini, ad esempio, hanno raggiunto e superato le performance di medici esperti nell'identificazione di varie patologie[896]; l'analisi dei dati con strumenti di intelligenza artificiale ha permesso lo sviluppo di accurati sistemi valutativi e predittivi[897], il cui coinvolgimento nella decisione medica, come si dirà, pone questioni etiche particolarmente complesse.

Il sistema sanitario, dunque, è uno dei campi di applicazione privilegiata delle tecnologie di intelligenza artificiale. Analisi autorevoli ritengono che, in futuro, l'integrazione di sistemi intelligenti nell'attività medica crescerà ulteriormente, specialmente se i regolatori del settore sapranno porre rimedio alle principali criticità sollevate da alcune ipotesi di utilizzo (ad

[893] Si rimanda, in primo, luogo alle informazioni pubblicate dallo stesso *PubMed*: https://bit.ly/3BgBG5H (6 agosto 2022). Cfr., inoltre, V. KAUL, S. ENSLIN, S.A. GROSS, *History of artificial intelligence in medicine*, in *Gastrointestinal endoscopy*, 10, 92, 4, 2020, p. 807-808.

[894] Cfr. *ex multis* T. LANE, *A short history of robotic surgery*, in *Annals of the Royal College of Surgeons of England*, 100, 6, 2018, p. 5-7 e A. BRODIE, N. VASDEV, *The future of robotic surgery*, in *Annals of the Royal College of Surgeons of England*, 100, 7, 2018, p. 4-13.

[895] Il riferimento obbligato è, in primo luogo, alla ricca e approfondita analisi di E. TOPOL, *Deep medicine. How artificial intelligence can make healthcare human again*, New York, 2019: cfr. inoltre V. KAUL, S. ENSLIN, S.A. GROSS, *History of artificial intelligence in medicine cit.*, p. 809 ss.; F. PICCIALLI, V. DI SOMMA, F. GIAMPAOLO, S. CUOMO, G. FORTINO, *A survey on deep learning in medicine: Why, how and when?*, in *Information Fusion*, 66, 2021, p. 111-137; F. WANG, L. CASALINO, D. KHULLAR, *Deep Learning in Medicine—Promise, Progress, and Challenges*, in *JAMA Internal Medicine*, 179, 3, 2019, p. 293 ss.; L. SCAFFARDI, *La medicina alla prova dell'intelligenza artificiale*, in *DPCE – Online*, 1, 2022, p. 349 ss.

[896] R. GARGEYA, T. LENG, *Automated Identification of Diabetic Retinopathy Using Deep Learning*, in *Ophthalmology*, 124, 7, 2017, p. 962-969; A. ESTEVA, B. KUPREL, R. A. NOVOA, J. KO, S. M. SWETTER, H. M. BLAU, *Dermatologist-level classification of skin cancer with deep neural networks*, in *Nature*, 542, 7639, 2017, p. 115-118; S. MATHOTARACHI, M. ZHU, C. XU, J. YU, Y. WU, C. LI, M. ZHANG, *Differentiation of Pancreatic Cancer and Chronic Pancreatitis Using Computer-Aided Diagnosis of Endoscopic Ultrasound (EUS) Images: A Diagnostic Test*, in *PLoS ONE*, 8, 5, 2013.

[897] S. F. WENG, J. REPS, J. KAI, J. M. GARIBALDI, N. QURESHI, *Can machine-learning improve cardiovascular risk prediction using routine clinical data?*, in *PLOS ONE*, 12, 4, 2017; T. A. PASCOAL, M. SHIN, A. L. BENEDET, M. KANG, T. BEAUDRY, *Identifying incipient dementia individuals using machine learning and amyloid imaging*, in *Neurobiology of Aging*, 59, 2017, p. 80-90.

esempio,riguardo al collocamento di eventuali responsabilità, o, come si dirà, al controllo sulla tecnologia e alla relazione col paziente)[898]. In ragione di questo scenario, cercare di sistematizzare leprincipali innovazioni utilizzate in ambito sanitario risulta ancor più complesso che in altri contesti, e ogni categorizzazione proposta si espone a inevitabili obiezioni di incompletezza. Fermo questo limite, al fine di individuare le applicazioni dell'intelligenza artificiale in campo medico più significative, devono almeno menzionarsi:

- **sistemi di chirurgia robotica**. L'automazione di operazioni chirurgiche è diventata, negli ultimi decenni, sempre più comune. L'utilizzo di sistemi di *computer vision*, sonde e braccia meccaniche permette di intervenire sul corpo del paziente in modo meno invasivo che con le tecniche di chirurgia tradizionale. L'addestramento all'utilizzo di sistemi avanzati, ormai, è parte a pieno titolo della formazione medica specialistica nei settori in cui il loro impiego è più diffuso, come nefrologia e urologia[899]. Le prime tecnologie di chirurgia robotica sono apparse sul mercato negli anni '80 e '90, con l'avvento di sistemi come PROBOT[900], per interventi di resezione della prostata, o ROBODOC[901], finalizzato a rendere meno invasive le operazioni interessanti l'articolazione coxo-femorale. A partire dal primo decennio degli anni 2000, invece, il settore è stato pressoché dominato dal *Da Vinci Surgical System*, strumento sviluppato dall'azienda americana *Intuitive Surgical*, composto da un numero variabile di braccia meccaniche e utilizzato per diverse operazioni[902]. Tutte queste tecnologie si basano sulla robotizzazione dell'intervento sul paziente, svolto dalla macchina con livelli variabili di autonomia, sotto il controllo, o l'integrale teledirezione, del sanitario, tipicamente – è il caso, ad esempio, del sistema *Da Vinci* - attraverso un apposito monitor. Includere o meno nella famiglia delle tecnologie di intelligenza artificiale i sistemi che operano sotto la completa direzione del medico dipende largamente dalla definizione di IA adottata. In ogni caso, numerosi studi recenti evidenziano le prospettive aperte, in un futuro prossimo, dall'integrazione in strumenti di

[898] Cfr. ad es. Deloitte, *The future of AI in healthcare. How AI will impact patients, clinicians and the pharmaceutical industry*, https://bit.ly/3Ql9qmM, 2019; I. KICKBUSCH, D. PISELLI, A. AGRAWAL, R. BALICER, O. BANNER, M. ADELHARDT ET AL., *The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world*, in *Lancet*, 398, 10312, 2021, p. 1727-1776; T. DAVENPORT, R. KALAKOTA, *The potential for artificial intelligence in healthcare*, in *Future Healthcare Journal*, 6, 2, 2019, p. 94-98.

[899] T. DAVENPORT, R. KALAKOTA, *The potential for artificial intelligence in healthcare cit.*, p. 95.

[900] S.J. HARRIS, F. ARAMBULA-COSIO, Q. MEI, R.D. HIBBERD, B.L. DAVIES, J.E.A. WICKHAM, *The Probot—an active robot for prostate resection*, in *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 211, 4, 1997, p. 317-325.

[901] T. LANE, *A short history of robotic surgery*, p. 6.

[902] Cfr. ad esempio C. FRESCHI,V. FERRARI, F. MELFI, M. FERRARI, F. MOSCA, A. CUSCHIERI, *Technical review of the da Vinci surgical telemanipulator*, in *The International Journal of Medical Robotics and Computer Assisted Surgery*, 9, 4, p. 396-406.

chirurgia robotica delle tecniche intelligenza artificiale più avanzate, in primo luogo l'apprendimento automatico per mezzo delle reti neurali[903].

- **Database sanitari ed elaborazione del linguaggio naturale**. La digitalizzazione di dati sanitari è cruciale per l'efficienza dell'assistenza medica al singolo paziente e dell'intero sistema sanitario. Infatti, la possibilità di consultare informazioni sui precedenti clinici del paziente è indispensabile per una corretta diagnosi e la definizione del miglior trattamento applicabile. Inoltre, come si dirà, maggiore è la disponibilità di dati per l'elaborazione, migliore è la capacità della ricerca di investigare possibili concause di patologie, identificare comportamenti a rischio, sviluppare nuove strategie di cura. Come già detto, la predisposizione di banche dati digitali e la raccolta di informazioni in formato elettronico ed interoperabile sono molto più comuni in ambito sanitario che in altri settori. Ciò nonostante, moltissime informazioni sanitarie (diagnosi, prescrizioni mediche, articoli e testi scientifici risalenti) sono, ancora oggi, diffuse in formati non digitali, in primo luogo la carta, o comunque difficilmente integrabili nei comuni *database*[904]. I sistemi di elaborazione del linguaggio naturale si stanno dimostrando molto efficaci per mitigare questo problema, permettendo di automatizzare l'elaborazione anche di tali documenti[905]. Inoltre, l'analisi dei database sanitari con tali strumenti, integrati con altre tecniche di intelligenza artificiale, consente di accedere a risultati di inedita qualità – ad esempio riguardo all'identificazione di casi simili a una determinata vicenda clinica – in modo analogo a quanto già analizzato in materia di motori di ricerca giurisprudenziali intelligenti[906].
- **Intelligenza artificiale impiegata nell'automazione di attività burocratiche e amministrative e nei rapporti col pubblico**. Anche in ambito sanitario, le possibilità di impiego di una pluralità di tecnologie basate in tutto o in parte sull'intelligenza artificiale sembrano notevoli. Tecniche di NLP potrebbero, ad esempio, essere utilizzate per la redazione di documenti, mentre sistemi gestionali basati su strumenti avanzati di analisi dei dati potrebbero rendere più efficiente l'organizzazione dei carichi di lavoro, portando, tra le altre cose, alla diminuzionedella lunghezza delle liste d'attesa. Il tema pare urgente, poiché studi a riguardo, condotti negli Stati Uniti, hanno rilevato che infermieri e infermiere

[903]Cfr. *ex multis* M. BHANDARI, T. ZEFFIRO, M. REDDIBOINA, *Artificial intelligence and robotic surgery: current perspective and future directions*, in *Current Opinion in Urology*, 30, 1, 2020, p. 48-54.
[904]Cfr. Deloitte, *The future of AI in healthcare*, p. 5.
[905]T. DAVENPORT, R. KALAKOTA, *The potential for artificial intelligence in healthcare cit.*, p. 95, ancora Deloitte, *The future of AI in healthcare*, p. 5.
[906]Si vedano, *ex multis*, le analisi di C. DREISBACH, T.A. KOLECK, S. BOURNE, S. BAKKEN, *A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data*, in *International Journal of Medical Informatics*, 125, 2019, p. 37-46; O. BACLIC, M. TUNIS, K. YOUNG, C. DOAN, H. SWERDFEGER, J. SCHONFELD, *Challenges and opportunities for public health made possible by advances in natural language processing*, in *Canada Communication Disease Report*, 4, 46, 6, p. 161-168.

trascorrono quasi un quarto del loro orario di lavoro impegnati in attività amministrative, e l'automazione permetterebbe di ridestinare alla cura dei pazienti almeno parte di questo tempo[907].Inoltre, alcune strutture sanitarie, specialmente nel mondo anglosassone, utilizzano *chatbot* per guidare gli utentinello svolgimento in autonomia di alcune attività, come la prenotazione *online* della visita specialistica più adatta ai sintomi da loro sofferti[908]. Alcune ricerche hanno evidenziato difficoltà di utilizzo e bassi livelli di soddisfazione da parte di diversi utenti[909]; non può non sottolinearsi, inoltre, come sembri necessaria una riflessione sui potenziali rischi connessi al*digital divide*, vista l'età tipicamente avanzata di molti utenti del sistema sanitario.

- **Machine learning ed elaborazione del linguaggio naturale per la ricerca farmaceutica**. Com'è noto, la ricerca in ambito farmacologico si basa su sperimentazioni divise in più fasi, e sono necessari diversi anni per l'identificazione del principio attivo più adatto a un determinato scopo terapeutico e la verifica dell'assenza di effetti avversi sull'essere umano. Recenti ricerche hanno messo in luce come tecnologie avanzate di intelligenza artificiale possano rendere più rapide ed efficienti entrambe le fasi del procedimento. In particolare, tecnologie basate sull'elaborazione del linguaggio naturale potrebbero accelerare la verifica della sostenibilità teorica di un'ipotesi di lavoro e riscontrare eventuali legami con ricerche passate, permettendo di analizzare, in pochi attimi, migliaia di pagine di manuali e studi scientifici [910] . Inoltre, determinati strumenti di *machine learning*si sono dimostrati particolarmente efficaci nell'elaborare modelli predittivi di quali molecole si dimostrino più efficaci per ottenere un determinato effetto biologico, permettendo, potenzialmente, di accelerare procedimenti di scelta che durano normalmente diversi anni[911].
- **Tecnologie basate sull'elaborazione di dati con sistemi di intelligenza artificiale con finalità decisionali o predittive**. Com'è noto, i recenti progressi nel campo dell'apprendimento automatico, in particolare per ciò che riguarda le reti neurali

[907]Cfr. J. COMMINS, *Nurses say distractions cut bedside time by 25%*, in *HelathLeader*, 2010, https://bit.ly/3RKwwog, (6 agsoto 2022).
[908]In particolare, il Regno Unito, come già visto, ha reso pubblica l'intenzione di investire fortemente, nei prossimi anni, sullo sviluppo di *chatbot* per incrementare l'efficienza del sistema sanitario, cfr. ad es. R. DAWS, *The NHS hopes an AI chatbot will help tackle patient wait times*, in *AI News*, 29 luglio 2022, https://bit.ly/3qgXRT5 (6 agosto 2022).
[909]Cfr. UserTesting, *healthcare chatbot apps are on the rise but the overall custode experience (cx) falls short according to a UserTesting report*, 2019, https://bit.ly/3ATovGM, (6 agosto 2022).
[910] R. MCENTIRE, D. SZALKOWSKI, J. BUTLER, M. S. KUO, M. CHANG ET AL, *Application of an automated natural language processing (NLP) workflow to enable federated search of external biomedical content in drug discovery and development*, in *Drug Discovery Today*, 21, 5, 2016, p. 826-835; H.ÖZTÜRK, A.ÖZGÜR, P.SCHWALLER, T.LAINO, E.OZKIRIMLI, *Exploring chemical space using natural language processing methodologies for drug discovery*, in *Drug Discovery Today*, 25, 4, 2020, p. 689-705.
[911]H.CHEN, O. ENGKVIST, Y. WANG, M. OLIVECRONA, T. BLASCHKE, *The rise of deep learning in drug discovery*, in *Drug Discovery Today*, vol. 23, 6, 2018, pp. 1241-1250; A.LAVECCHIA, *Deep learning in drug discovery: opportunities, challenges and future prospects*, in *Drug Discovery Today*, 24, 10, 2019, p. 2017-2032.

profonde,hanno generato un picco d'interesse per il possibile impiego di tale famiglia di tecnologie per l'elaborazione di modelli valutativi e predittivi. Il settore sanitario è uno dei campi applicativi d'elezione di questa rivoluzione tecnologica, per ragioni in parte già citate[912]. In primo luogo, l'importanza degli interessi coinvolti e la solidità dei potenziali finanziatori – sistemi sanitari pubblici e multinazionali operanti in ambito medico e farmaceutico – sembrano garantire margini di guadagno elevati per gli sviluppatori di tecnologie efficaci da impiegare nell'ambito. In secondo luogo, il settore è caratterizzato da una forte apertura alla tecnologia: sono disponibili *database* digitalizzati di notevoli dimensioni, adatti all'elaborazione[913], e la crescente diffusione di strumenti avanzati – si pensi ai c.d. dispositivi *wearable* - fa apparire la raccoltà di nuovi dati particolarmente agevole[914]. Gli ultimi anni sono stati caratterizzati dal proliferare di sistemi basati sull'elaborazione con tecniche di intelligenza artificiale di dati attinenti all'ambito sanitario, al fine di sviluppare modelli predittivi o decisionali. Alcuni di questi hanno già assunto un ruolo di rilievo sul mercato, mentre altri, allo stato dell'arte, rappresentano semplici ipotesi di ricerca.In alcuni casi, le valutazioni frutto dell'analisi dei dati sono lo scopo ultimo del sistema, mentre in altri casi sono impiegate per giungere a un risultato diverso. È possibile ipotizzare una suddivisione di questa famiglia di tecnologie in tre categorie principali: i sistemi che automatizzano l'attività diagnostica dei medici umani; i sistemi di personalizzazione dell'attività di assistenza al paziente; i sistemi impiegati per strategie di pianificazione sanitaria. Come in altre circostanze, le sovrapposizioni non mancano e si tratta di categorie meramente tendenziali.

Per quanto riguarda le tecnologie basate sull'apprendimento automatico impiegate per la diagnosi, devono menzionarsi, in particolare, alcuni sistemi di *image recognition*. L'utilizzo di reti neurali profonde in tale ambito ha permesso di raggiungere prestazioni accurate

[912] Si rimanda ancora, in primo luogo, a E. TOPOL, *Deep medicine cit.*; cfr. anche Deloitte, *The future of AI in healthcare cit*; I. KICKBUSCH, D. PISELLI, A. AGRAWAL, R. BALICER, O. BANNER, M. ADELHARDT ET AL., *The Lancet and Financial Times Commission on governing health futures 2030 cit.*; T. DAVENPORT, R. KALAKOTA, *The potential for artificial intelligence in healthcare cit.*; V. KAUL, S. ENSLIN, S.A. GROSS, *History of artificial intelligence in medicine cit.*; F. PICCIALLI, V. DI SOMMA, F. GIAMPAOLO, S. CUOMO, G. FORTINO, *A survey on deep learning in medicine cit*.; F. WANG, L. CASALINO, D. KHULLAR, *Deep Learning in Medicine cit*.

[913] Oltre ad esempi come il già citato *PubMed*, l'importanza e la ricchezza dei database medici esistenti, e la necessità di incrementarne accuratezza e dimensioni, parallelamente a quella di garantire il rispetto dei principi ispiratori di un corretto trattamento dei dati personali, è dimostrata dalla volontà di costruire, sullo scenario europeo, uno spazio comune dei dati sanitari, in merito al quale, come già analizzato, è stata di recente presentata una proposta di Regolamento della Commissione, cfr. COMMISSIONE EUROPEA, *Proposta di Regolamento del Parlamento Europeo e del Consiglio sullo spazio europeo dei dati sanitari*, COM(2022) 197 final.

[914] V. VIJAYAN, J. P. CONNOLLY, J. CONDELL, N. MCKELVEY, P. GARDINER, *Review of Wearable Devices and Data Collection Considerations for Connected Health*, in *Sensors*, 21, 16, 2021, p. 5589 ss.; J. GASKIN, J. JENKINS, T. MESERVY, J. STEFFEN, K. PAYNE, *Using Wearable Devices for Non-invasive, Inexpensive Physiological Data Collection*, in *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017, http://hdl.handle.net/10125/41221 (8 agosto 2022).

quanto quelle di sanitari umani esperti – e, talvolta, di superare questi ultimi – per la diagnosi di alcune patologie. Hanno acquisito notorietà, in particolare, alcuni sistemi in grado di operare diagnosi accurate, e particolarmente precoci, di forme di melanoma analizzando l'immagine del tessuto interessato, il cui successo fa apparire possibile lo sviluppo di strumenti di analoga efficacia anche per altre patologie tumorali[915]. Risultati notevoli sono stati raggiunti anche nella diagnosi di alcune tipologie di retinopatia e per la distinzione tra prostatite e i primi stadi del tumore alla prostata[916]. L'analisi di grandi database di dati sanitari con tecnologie di intelligenza artificiale, inoltre, sembra poter facilitare la diagnosi di malattie rare, la cui eziologia è spesso di difficile identificazione[917]. Allo stato dell'arte, però, si tratta di un'ipotesi non ancora consolidata.

I sistemi di intelligenza artificiale impiegati per la personalizzazione dell'assistenza al paziente e il monitoraggio delle proprie condizioni di salute, invece, sono una famiglia di tecnologie ampia e variegata. Hanno già conquistato un ruolo di primo piano sul mercato varie tipologie di assistenti sanitari digitali, che utilizzano dati relativi alla situazione psicofisica dell'utente, spesso raccolti con dispositivi indossabili (i c.d. *wearable*, già menzionati) per elaborare valutazioni e predizioni sulla sua condizione sanitaria, consigliare stili di vita più corretti, sollecitare controlli medici o ipotizzare la necessità di terapie[918]. Il settore è in rapida espansione, con applicazioni di varia natura[919]; per limitarsi a un esempio particolarmente noto ed avanzato, può citarsi il*chatbot*implementato, a partire dal 2018, dall'azienda britannica di *e-health* Babylon a supporto del sistema sanitario ruandese[920].

[915] Cfr. A. HEKLER ET AL., *Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images*, in *European Journal of Cancer*, 118, 2019, p. 91 ss.; A. ESTEVA, B. KUPREL, R. A. NOVOA, J. KO, S. M. SWETTER, H. M. BLAU, *Dermatologist-level classification of skin cancer with deep neural networks cit.*Per quanto riguarda le prospettive di sviluppi ulteriori in ambito oncologico, cfr. ad es. Z. HU, J. TANG, Z. WANG, K. ZHANG, L. ZHANG, Q. SUN, *Deep learning for image-based cancer detection and diagnosis − A survey*, in *Pattern recognition*, 83, 2018, p. 134-149; D.A. Alkurdi, M. & J. Ilyas, A. Jamil, *Cancer detection using deep learning techniques* in *Evolutionary Intelligence*, 2021, https://doi.org/10.1007/s12065-021-00635-5 (8 agosto 2022).

[916] Cfr. S. MATHOTARACHI, M. ZHU, C. XU, J. YU, Y. WU, C. LI, M. ZHANG*, Differentiation of pancreatic cancer and chronic pancreatitis using computer-aided diagnosis of endoscopic ultrasound (EUS) images cit.*;R.GARGEYA, T. LENG,*Automated identification of diabetic retinopathy using deep learning cit.*

[917] Cfr. ad esempio S. BRASIL, C. PASCOAL, R. FRANCISCO, V. A. DOS REIS FERREIRA, P. VIDEIRA, G. VALADÃO, *Artificial Intelligence (AI) in Rare Diseases: Is the Future Brighter?*, in *Genes*, 10, 12, 2019, p. 978 ss.; J. LEE, C. LIU, J. KIM, Z. CHEN, Y. SUN, J.R. ROGERS ET AL., *Deep learning for rare disease: a scoping review*, medRxiv, 2022, https://doi.org/10.1101/2022.06.29.22277046 (8 agosto 2022).

[918] Per una panoramica sui dispositivi pià utilizzati cfr. J. LOUCKS, D. STEWART, A. BUCAILLE, G. CROSSAN, *Wearable technology in health care: getting better all the time*, Deloitte, 1 dicembre 2021, https://bit.ly/3QnEm64 (8 agosto 2022); K. CURTIS, *Wearable tech in healthcare: top devices making a difference*, in Edumed, https://bit.ly/3RwWiwv (8 agosto 2022). Peraltro, alcune di tali applicazioni, ad esempio quelle volte al monitoraggio della frequenza cardiaca, possono avere un effetto salvavita, cfr. R. HUTTER EPSTEIN*, Can a smartwatch save your life?*, The New York Times, 26 luglio 2021.

[919] Per la situazione attuale e le principali prospettive di ricerca future cfr. L. LU, J. ZHANG, Y. XIE, F. GAO, S. XU, X. WU ET AL., *Wearable health devices in health care: narrative systematic review*, in *JMIR mHealth and uHealth*, 8, 11, 2020, https://doi.org/10.2196/18907 (8 agosto 2022).

[920]J. BIZIMUNGU, *Babyl's chatbot to enhance digital healthcare platform*, The New Times, 11 gennaio 2018.

L'applicazione permette ai pazienti di esporre sintomi e altri dubbi relativi alla salute, in formato testuale o con messaggi vocali, e, con tecnologie di machine learning, formula diagnosi e indica trattamenti. Lo strumento, inoltre, prevede la possibilità di accedere ai servizi di telemedicina che l'azienda da tempo offre in vari paesi del mondo, rendendo così possibile, qualora il *chatbot* lo reputi opportuno, interagire a distanza con personale sanitario umano[921]. Nonostante la tecnologia abbia di certo migliorato le possibilità di accesso all'assistenza sanitaria di larghe fasce della popolazione, deve segnalarsi che la correttezza delle valutazioni cliniche del *chatbot* è stata, talvolta, messa in discussione[922].

Tra i sistemi impiegati per personalizzare l'assistenza al paziente vanno inclusi anche alcuni algoritmi teorizzati di recente, in grado di prevedere il decorso clinico del paziente in modo più preciso di quanto permettano le valutazioni del personale sanitario[923]. Si tratta, chiaramente, di innovazioni tecnologiche connesse alla possibilità di sviluppare, attraverso l'analisi di dati sanitari con reti neurali profonde, modelli predittivi particolarmente accurati. La diffusione di tali strumenti permetterebbe di individualizzare il percorso di cura, incrementandone in modo decisivo l'efficacia, e, per quanto riguarda i pazienti ospedalizzati, consentirebbe un'allocazione più efficiente delle risorse sanitarie e un incremento generale della qualità delle prestazioni[924]. Sarebbe possibile, infatti, valutare con precisione tipologia e durata dell'assistenza necessaria a ciascun degente e, di conseguenza, organizzare in modo più efficiente ed economico l'intera struttura ospedaliera. Come si dirà, la ricerca dimostra estremo interesse verso queste applicazioni dell'intelligenza artificiale, mentre non è semplice reperire informazioni sul loro utilizzo nella pratica clinica. Alcune di esse, affrontate nel dettaglio in seguito, sollevano delicatissime questioni etiche, poiché la possibilità di formulare predizioni accurate

[921] Cfr. la pagina in proposito sul sito web di *Babylon Health*: https://www.babylonhealth.com/en-us/babyl (8 agosto 2022).

[922] In particolare, uno studio, pubblicato nel 2018 su *Lancet*, ha messo in evidenza come l'affermazione dell'azienda sviluppatrice Babylon Health, secondo cui il *chatbot* garantiva prestazioni equivalenti o superiori a quelle dei medici umani, non risultasse supportata da evidenze scientifiche solide, cfr.H. FRASER, E. COIERA, D. WONG, *Safety of patient-facing digital symptom checkers*, in *The Lancet*, 392, 10161, 2018, p. 2263-2264. Si vedano, inoltre, le considerazioni di C. CASONATO, *Intelligenza artificiale e medicina: l'impatto sulla relazione di cura (cenni)*, in U. SALNITRO (a cura di), *SMART – La persona e l'infosfera. Atti del Convegno 30 settembre - 2 ottobre 2021Catania*, Pisa, 2022, p. 107 ss.

[923] Tra i numerosissimi pubblicati, con cadenza ormai quasi quotidiana, nelle principali riviste mediche cfr. G. V. GLINSKY, A. B. GLINSKII, A. J. STEPHENSON, R. M. HOFFMAN, W. L. GERALD, *Gene expression profiling predicts clinical outcome of prostate cancer*, in *Journal of Clinical Investigation*, 113, 6, 2004, p. 913-923; D. ZHANG, L. ZOU, X. ZHOU, F. HE, *Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer*, in *IEEE Access*, 6, 2018, p. 28936 ss.; J. KWON, J. JEON, H.M. KIM ET AL., *Deep-learning-based out-of-hospital cardiac arrest prognostic system to predict clinical outcomes*, in *Resuscitation*, 139, 2019, p. 84-91.

[924] Cfr. in particolare K. YU, Z. YANG, C. WU, Y. HUANG, X. XIE,*In-hospital resource utilization prediction from electronic medical records with deep learning*, in *Knowledge-Based Systems*, 223, 2021, p. 107052 ss.

sull'evoluzione delle condizioni del paziente porta con sé, ovviamente, anche l'ipotesi di prevederne l'eventuale decorso clinico infausto[925].

Infine, sistemi predittivi possono essere utilizzati a supporto di decisioni di pianificazione sanitaria[926]. L'analisi di banche di dati sanitari relativi alla popolazione generale – si pensi ai dati stratificati, nel corso degli anni, nei fascicoli sanitari elettronici dei cittadini[927] – potrebbe permettere di identificare macrotendenze a cui rispondere con politiche adeguate. Può farsi l'ovvio esempio dell'invecchiamento della popolazione in molti paesi occidentali, dovuto all'allungarsi dell'aspettativa di vita e al calo del tasso di natalità, cui dovranno corrispondere idonee strategie di ristrutturazione del sistema sanitario[928]. La previsione, in questo caso, è di certo accessibile anche senza l'ausilio dell'intelligenza artificiale, ma l'analisi di grandi moli di dati permetterebbe valutazioni molto più complesse, come l'eventuale sovradiffusione di determinate patologie in alcune zone o fasce d'età, o tra soggetti comunque accomunati da alcune caratteristiche. Le autorità sanitarie potrebbero, a partire da tali conoscenze, sviluppare politiche di prevenzione mirate ed efficaci.

La combinazione di sistemi predittivi di questo genere con gli strumenti visti in precedenza, inoltre, permetterebbe lo sviluppo di strategie di c.d. medicina di iniziativa particolarmente efficaci, un'ipotesi su cui si sta concentrando sempre di più l'attenzione degli operatori sanitari, anche nel nostro Paese[929]. L'incrocio tra i dati sanitari del singolo paziente e i risultati di studi su database relativi alla popolazione generale, infatti, potrebbe permettere

[925] La circostanza ha riguardato, come si approfondirà *infra*, p. 293 ss., anche la pandemia di Covid-19, durante la quale sono stati teorizzati diversi algoritmi con cui prevedere le probabilità di peggioramento, e conseguente ricovero in terapia intensiva, dei pazienti, a breve e medio termine, cfr. ad esempio D. PATEL ET AL., *Machine learning based predictors for COVID-19 disease severity*, in *Scientific Reports*, 11, 2021, p. 4673; D. S. CHOW ET AL., *Development and external validation of a prognostic tool for COVID-19 critical disease*, in *PLoS One*, 15, 12, 2020.

[926] Cfr. ad esempio M.J. KHOURY, J. P. A. IOANNIDIS, *Big data meets public health*, in *Science*, 346, 6213, 2014, p. 1054 ss.; D. VALLE-CRUZ, E. A. RUVALCABA-GOMEZ, R. SANDOVAL-ALMAZAN, J. IGNACIO CRIADO, *A Review of Artificial Intelligence in Government and its Potential from a Public Policy Perspective*, in *Proceedings of the 20th Annual International Conference on Digital Government Research*, Dubai United Arab Emirates, 2019, p. 91-99, https://dl.acm.org/doi/10.1145/3325112.3325242 (9 agosto 2022); H. ASHRAFIAN, A. DARZI, *Transforming health policy through machine learning*, in *PLOS Medicine*, 15, 11, 2018, https://doi.org/10.1371/journal.pmed.1002692 (9 agosto 2022).

[927] Non possono non menzionarsi, sul punto, le questioni poste in materia di protezione dei dati personali da iniziative di questo tipo, sulle quali la letteratura giuridica di settore, del resto, ha da tempo concentrato i propri sforzi per lo sviluppo di architetture informatiche e giuridiche in grado di assicurare il più alto livello di rispetto della normativa in materia di dati personali, cfr. *ex multis* M.MOSTERT, A.L. BREDENOORD, M. C. I. H. BIESAART, J. J. M. VAN DELDEN, *Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach*, in*European Journal of Human Genetics*, 24, 7, 2016, p. 956-960; G. PASCUZZI, R. DUCATO, Biobanche *di ricerca tra proprietà, privacy e proprietà intellettuale: un approccio LawTech*, in*Notizie di Politeia*, 2012, p. 55.

[928] Sul tema, cfr. ad esempio L. P. FRIED, P. PACCAUD, *Editorial: The Public Health Needs for an Ageing Society*, in *Public Health Reviews*, 32, 2, 2010, p. 351-355; M. NOALE, F. LIMONGI, E. SCAFATO, S. MAGGI, G. CREPALDI, *Longevity and health expectancy in an ageing society: implications for public health in Italy*, in *Annali dell'Istituto Superiore di Sanità*, 48, 3, 2012, p. 292-299, DOI: 10.4415/ANN_12_03_10.

[929] Si veda, ad esempio, il *Patto per la salute 2019-2021* della Conferenza Stato-Regioni per la pianificazione delle attività del Sistema Sanitario Nazionale, che attribuisce un ruolo di primo piano alla medicina d'iniziativa, cfr. https://bit.ly/3RQBKyJ (9 agosto 2022).

l'identificazione precoce di fattori di rischio individualee l'implementazione di appositi programmi di monitoraggio e prevenzione[930]. Risulterebbe, così, realizzata l'ispirazione all'intervento preventivo, finalizzato al mantenimento di condizioni ottimali di salute e non limitato al momento della malattia, che ispira la medicina d'iniziativa.

3.2 La disciplina dell'intelligenza artificiale in medicina: l'impatto sui diritti fondamentali, l'approccio adottato dalla Proposta di Regolamento europeo sull'IA e una possibile indicazione proveniente dagli ordinamenti anglosassoni

L'ambito sanitario è, in generale, uno dei banchi di prova privilegiati del diritto. I beni coinvolti, a cominciare dalla salute, l'integrità fisica e la vita stessa, rendono particolarmente frequente il contenzioso nel campo. Inoltre, la solidità patrimoniale di medici, ospedali e case farmaceutiche non è di certo un deterrente alle azioni risarcitorie avverso di essi, tanto che la c.d. medicina difensiva – controlli e terapie non necessari, prescritti al fine di non incorrere in responsabilità – è considerata uno dei problemi principali di molti sistemi sanitari avanzati, Italia compresa[931]. Non è un caso, allora, che sia proprio l'ambito medico uno dei terreni di evoluzione di alcuni basilari istituti giuridici, a cominciare dalla responsabilità civile e penale (si pensi alle vicende giurisprudenziali, derivanti dall'attività medica,in materia di c.d. contatto sociale[932] o per il nesso di causalità penale)[933]. Allo stesso modo, gli strumenti utilizzati nell'attività medica sono soggetti a

[930] Sempre sullo scenario italiano, il tema è stato, peraltro, di recente oggetto di un parere del Garante per la Protezione dei Dati Personali, che ha fissato i requisiti che l'applicazione di tecnologie di intelligenza artificiale in attività di medicina d'iniziativa deve rispettare per risultare conforme alla normativa in materia di protezione dei dati personali, cfr. GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Parere alla Provincia autonoma di Trento su uno schema di regolamento concernente la medicina di iniziativa nel servizio sanitario provinciale*, 1 ottobre 2020, https://bit.ly/3qrl5pM (9 agosto 2022).

[931] Cfr. ad esempio G. PANTING, *Doctors on the defensive*, The Guardian, 1 aprile 2005, https://www.theguardian.com/society/2005/apr/01/health.comment (4 ottobre 2022); D. M. STUDDERT ET AL., *Defensive Medicine Among High-Risk Specialist Physicians in a Volatile Malpractice Environment*, in *JAMA*, 293, 21, 2005, p. 2609-2617. Per quanto riguarda l'Italia, com'è noto le dimensioni del fenomeno – e i costi per il sistema sanitario che ne derivano – sono state una delle ragioni principali che hanno portato il Legislatore a intervenire più volte sul'assetto della responsabilità sanitaria, prima con il D.L. n. 158 del 13 settembre 2012 (c.d. *decreto Balduzzi*), poi con la L. n. 24 dell'8 marzo 2017 (c.d. *legge Gelli-Bianco*), nel tentativo di limitare il contenzioso temerario contro il personale sanitario. Cfr. *Medicina difensiva. Ci costa 10 mld l'anno. La pratica almeno una volta al mese quasi l'80% dei medici. Il report del Ministero della Salute*, in *Quotidiano sanità*, 26 marzo 2015, https://bit.ly/3RsuYi0 (4 ottobre 2022); U. GENOVESE, R. ZOJA, A. FERRARIO, A. SERPETTI, P. MARIOTTI, *La medicina difensiva. Questioni giuridiche, assicurative, medico-legali*, Bologna, 2011.

[932] Cfr. G. RONGA, *Le varie ipotesi di responsabilità cosiddetta da "contatto sociale"*, in L. VIOLA (A CURA DI), *La responsabilità civile ed il danno*, Milano, 2007, 1, p. 90 ss.; S. ROSSI, *Contatto sociale (fonte di obbligazione)*, in *Digesto delle discipline privatistiche*, sez. civile, Appendice di aggiornamento V, Torino, 2010, p. 346 ss. In ogni caso, le riforme *Balduzzi* e *Gelli-Bianco*, citate alla nota precedente, hanno definitivamente superato la tesi del contatto sociale per quanto riguarda la responsabilità del professionista sanitario, che, in mancanza di un'apposita stipulazione col paziente, risponde *ex lege* in via aquiliana.

[933] Riguardavano l'attività sanitaria ad esempio, i fatti oggetto della nota pronuncia Cass., sez. un., sent. n. 30328 del 10 luglio 2002, *Franzese*, cui si deve l'impianto della teoria della causalità prevalentemente utilizzata, ancora oggi, dalle corti. Sul tema, in letteratura cfr. ad es. R. BARTOLI, *Il problema della causalità penale. Dai modelli unitari al modello differenziato*, Torino, 2010. Si rinvia, inoltre, ai fondamentali studi di F. STELLA, *Leggi scientifiche e spiegazione causale nel diritto penale*, 2 ed., Milano, 2000.

procedure di verifica e certificazione particolarmente sofisticate, ed è nota la complessità dei protocolli imposti per l'adozione di trattamenti innovativi o la sperimentazione di farmaci. L'avvento dell'intelligenza artificiale si inserisce in questo contesto di complessità giuridica. Le innovazioni ad essa connesse rappresentano un ulteriore elemento di tensione per le categorie giuridiche che l'attività medica mette in discussione: ad esempio, come già estensivamente commentato in dottrina, l'uso di strumenti caratterizzati da un elevato grado di autonomia solleva problemi notevoli per la configurazione tradizionale della responsabilità sanitaria[934].

L'impatto è importante e variegato anche dal punto di vista dei diritti fondamentali. La protezione della vita e della salute, ovviamente, viene in gioco, come sempre di fronte all'arte medica. Ma ogni genere di applicazione dell'intelligenza artificiale in campo sanitario pone sfide spinose per i diritti, forse più che in ogni altro ambito. L'elaborazione con strumenti avanzati di dati sanitari chiama in causa, infatti, il diritto fondamentale alla riservatezza; il loro utilizzo per finalità diagnostiche e predittivela libertà di coscienza, l'autodeterminazione eil principio di eguaglianza; l'ingresso nel rapporto medico-paziente dell'elemento tecnologico ha un impatto, in ultima analisi, sulla dignità umana nei suoi vari profili, in primo luogo dal punto di vista del consenso effettivo ed informato al trattamento. Un'analisi di ogni profilo potenzialmente attinente alle ripercussioni sui diritti fondamentali delle sempre più numerose tecnologie intelligenti utilizzate in ambito medico mancherebbe di coerenza e raggiungerebbe, probabilmente, dimensioni insostenibili per questo lavoro. I prossimi paragrafi, dunque, tratteranno nel dettaglio un solo ambito tra i molti appena elencati: le applicazioni dell'intelligenza artificiale a supporto della decisione medica, in analogia alle analisi già svolte della decisione amministrativa e giudiziaria. In particolare, verrano prese in esame le c.d. scelte tragiche[935], riguardanti l'allocazione di risorse sanitarie scarse, un ambito che, per l'importanza dei valori coinvolti, pare particolarmente adatto a una riflessione basata sui diritti come quella condotta in questo lavoro.

Prima di passare all'esame dei risvolti giuridici dei dilemmi tragici è opportuno, però, dare conto più nel dettaglio del panorama normativo riguardante le applicazioni dell'intelligenza artificiale in ambito sanitario. Come già detto, tali tecnologie portano all'intersezione di testi legislativi corposi, già di per sé in grado di porre questioni giuridiche di difficile soluzione: è il caso, ad esempio, delle già citate normative in materia di protezione dei dati, validazione e certificazione dei dispositivi

[934] Cfr. *ex multis* U. RUFFOLO, *L'Intelligenza artificiale in sanità: dispositivi medici, responsabilità e "potenziamento",* in *Giurisprudenza italiana*, 2,2021, p. 502-508; A. PERIN, *Standardizzazione, automazione e responsabilità medica. Dalle recenti riforme alla definizione di un modello d'imputazione solidaristico e liberale*, in *BioLaw Journal - Rivista di BioDiritto*, 1, 2019, p. 29 ss.; W. N. PRICE II, S. GERKE, I. G. COHEN, *Potential Liability for Physicians Using Artificial Intelligence*, in*JAMA*, 322, 18, 2019, p. 1765-1766; F. MOLNÁR-GÁBOR, *Artificial Intelligence in Healthcare: Doctors, Patients and Liabilities*, in T. WISCHMEYER, T. RADEMACHER (A CURA DI),*Regulating Artificial Intelligence*, Cham, 2020, https://doi.org/10.1007/978-3-030-32361-5_15 (5 settembre 2022).

[935] Il riferimento è, in primo luogo, a G. CALABRESI, P. BOBBITT, *Tragic Choices*, analizzato più in profondità ai paragrafi successivi.

medici e sperimentazione sull'essere umano. L'ambito medico, dunque, è regolato da complesse discipline di settore, volte ad assicurare i migliori standard di sicurezza, al pari di altre attività connotate da un'elevata utilità sociale e alti livelli di rischio. Nonostante ciò, la Proposta di Regolamento delle Commissione Europea sull'intelligenza artificiale non sembra riconoscere questa specificità. La circostanza, del resto, pare una delle conseguenze dell'approccio d'ampio respiro adottato dal testo legislativo, che ha tra le sue principali finalità dettare una disciplina globale, applicabile a ogni sistema di intelligenza artificiale. La salute, in ogni caso, è uno dei beni primari che sono stati presi in considerazione, secondo quanto riporta il *Memorandum* esplicativo che accompagna la Proposta,per definire le classi di rischio delle tecnologie intelligenti, assieme alla sicurezza e alla protezione dei diritti fondamentali[936].

Al di là di questa petizione di principio, però, il testo normativo non contiene disposizioni specificamente rivolte all'applicazione dell'intelligenza artificiale in ambito sanitario, la cui disciplina, allora, dipende unicamente dalla classe di rischio di appartenenza. Da questo punto di vista, pare potersi affermare che la quasi totalità delle tecnologie intelligenti utilizzate in campo medico sia destinata, in caso di effettiva approvazione della Proposta, a rientrare tra i sistemi ad alto rischio. L'art. 6, infatti, com'è noto, comprende nella categoria i sistemi sottoposti alla disciplina del c.d. *new legislative framework* la cui immissione sul mercato o messa in servizio è subordinata al controllo da parte di un soggetto terzo[937]. La disciplina di matrice europea in materia di dispositivi medici (parte, per l'appunto, del *new legislative framework*)[938] impone il coinvolgimento di organismi terzi nella validazione e certificazione della grande maggioranza degli strumenti, esentando solamente i dispositivi più semplici, cui corrisponda un livello di rischio estremamente basso. La complessità dei sistemi di intelligenza artificiale tipicamente impiegati in ambito sanitario

[936]Il *Memorandum*, infatti, riporta, alle pagg. 3-4: «La proposta fissa regole armonizzate per lo sviluppo, l'immissione sul mercato e l'utilizzo di sistemi di IA nell'Unione seguendo un approccio proporzionato basato sul rischio. Essa propone un'unica definizione di IA adeguata alle esigenze future. Talune pratiche di IA particolarmente dannose sono vietate in quanto in contrasto con i valori dell'Unione, mentre sono proposte restrizioni e tutele specifiche in relazione a determinati usi dei sistemi di identificazione biometrica remota a fini di attività di contrasto. La proposta stabilisce una solida metodologia per la gestione dei rischi impiegata per definire i sistemi di IA "ad alto rischio" che pongono rischi significativi per la salute e la sicurezza o per i diritti fondamentali».

[937]Come già evidenziato, il sistema di rinvii agli allegati del testo legislativo con cui l'art. 6 della Proposta di Regolamento definisce il perimetro dei sistemi ad alto rischio risulta di non facile lettura. Sono applicazioni ad alto rischio tutti i sistemi di intelligenza artificiale che siano prodotti o componenti di sicurezza di prodotti disciplinati dalla normativa di armonizzazione indicata all'allegato II (che include, per l'appunto, le discipline del c.d. *new legislative framework*), per i quali sia prevista una valutazione di conformità da parte di terzi. Oltre a queste, sono applicazioni ad alto rischio anche quelle indicate all'allegato III, individuate per la delicatezza del contesto di utilizzo e delle finalità per cui sono impiegate (l'elenco comprende, ad esempio, i sistemi utilizzati nell'attività creditizia, per la gestione dei flussi migratori, o per l'identificazione biometrica nei casi in cui il precedente art. 5 la permette). Cfr. *supra*, p. 69 ss.

[938]Il c.d. *new legislative framework* è un pacchetto di misure di armonizzazione introdotte a livello europeo a partire dal 2008, al fine di migliorare l'efficacia dei controlli di conformità su determinati prodotti immessi nel mercato unico. Cfr. più ampiamente la panoramica fornita nella sezione *Internal market, industry, entrepreneurship and SME* del sito web della Commissione Europea: https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en (8 settembre 2022).

fa sì che essi ben difficilmente rientranno tra questi ultimi, e, dunque, ai fini della Proposta di Regolamento in materia di intelligenza artificiale saranno classificati come tecnologie ad alto rischio[939]. Essi dovranno rispettare l'ampio elenco di requisiti di sicurezza previsto per tale categoria di applicazioni, già analizzato in più parti di questo lavoro. Un insieme di accorgimenti tecnici di certo adatto ad assicurare il mantenimento di elevati standard di sicurezza e il rispetto dei diritti primari coinvolti in ambito sanitario, ma che, come si vedrà anche più avanti trattando dell'uso di algoritmi nelle scelte tragiche, a volte non sembra rispettosodelle specificità del campo.

Del resto, il comportamento degli ordinamenti extraeuropei sembra confermare che il settore sanitario presenta peculiarità importanti, delle quali, forse, sarebbe bene tenere conto al momento di regolare l'utilizzo dell'intelligenza artificiale all'interno di esso. Da questo punto di vista, pare significativo, in particolare, che le autorità di tre ordinamenti che si sono, finora, dimostrati tendenzialmente più liberali di quello europeo in materia di regolazione delle tecnologie – banalmente, non vi si registrano proposte legislative d'alto livello, allo stato dell'arte, per disciplinare l'intelligenza artificiale nel suo complesso[940] – abbiano di recente adottato una policy congiunta in materia di tecnologie intelligenti in ambito medico. Nell'ottobre del 2021, infatti, la *Food and Drug Administration* americana, il ministero canadese della sanità e la *Medicine & Healthcare products Regulatory Agency* britannica hanno emanato un elenco di dieci linee guida intitolato *Good Machine Learning Practice for Medical Device Development: Guiding Principles* [941], al fine di orientare gli operatori del mercato verso un corretto utilizzo dell'apprendimento automatico nello sviluppo di dispositivi medici. Il documento, ad ogni modo, è estremamente sintetico e generale e gli accorgimenti da esso previsti, sullo scenario europeo, risulterebbero garantiti dal combinato tra la Proposta di Regolamento sull'intelligenza artificiale e la normativa in materia di certificazione dei dispositivi medici già esistente[942]. È di certo, però, degno

[939] Com'è noto, la disciplina eurounitaria in materia di dispositivi medici, di recente innovata dall'entrata in vigore del Regolamento UE 2017/745, distingue i dispositivi in quattro classi differenti: I, IIa, IIb, III. Solo la certificazione dei dispositivi di classe I è svolta in autonomia dal fabbricante, mentre per quanto riguarda i dispositivi delle altre classi è sempre previsto l'intervento di un organismo esterno. È l'allegato VIII del Regolamento a stabilire i criteri di suddivisione dei dispositivi nelle varie classi. Il ruolo della classe I è estremamente ridotto e limitato agli strumenti più semplici (è il caso, ad esempio, delle lampade per l'illuminazione del paziente, o di garze e cerotti). La larga parte dei sistemi di intelligenza artificiale impiegati nell'attività medica appartiene, come già detto, alle altre classi, per la cui certificazione è disposto l'intervento di terzi, e il Regolamento UE 2017/245 rientra nelle norme indicate all'Allegato II della Proposta di Regolamento in materia di intelligenza artificiale. Ai sensi della disciplina dell'art. 6 della Proposta riassunta alla nota precedente, allora, essi rientreranno tra le applicazioni dell'intelligenza artificiale ad alto rischio.

[940] Con la specifica che nell'ordinamento canadese è comunque presente la *Directive on automathed decision making*, analizzata in più punti del lavoro, che ha comunque un campo di applicazione relativamente ampio (i processi decisionali della Pubblica Amministrazione).

[941] Il testo del documento è consultabile su sito dell'FDA americana: https://bit.ly/3RFIBKU (8 settembre 2022).

[942]Il documento individua 10 *guiding principles*: Multi-disciplinary expertise is leveraged throughout the total product life cycle; Good software engineering and security practices are implemented; Clinical study participants and data sets are representative of the intended patient population; Training data sets are independent of test sets; Selected reference datasets are based upon best available methods; Model design is tailored to the available data and reflects the intended use of the device; Focus is placed on the performance of the human-AI team; Testing demonstrates device performance

di nota che, allo stato dell'arte, non si rivengano policy o proposte normative analoghe sullo scenario europeo, che riconoscano le specificità del campo sanitario anche dal punto di vista della disciplina dell'IA, nonostante la maggiore attenzione alla regolazione dello sviluppo tecnologico che caratterizza l'ordinamento dell'Unione.

3.3. Le scelte tragiche: definizione, caratteristiche e applicabilità del concetto all'ambito sanitario

La tragedia della Grecia antica rappresentava frequentemente personaggi alle prese con scelte irrisolvibili, scaturenti dal conflitto tra valori considerati non negoziabili. È questo il significato essenziale della parola *dilemma*, etimologicamente, per l'appunto, "preposizione doppia"[943]. Proprio il dilemma, e la possibilità di risolverlo solamente compiendo un sacrificio che appare moralmente ingiusto, connota le decisioni che siamo abituati a definire "tragiche".

A ben vedere, il riferimento allo scontro tra principi non è sufficiente a definire con precisione il campo delle scelte tragiche. Resta sullo sfondo, infatti, il problema di quali siano tali valori fondamentali, e in quali casi non esista una gerarchia che permetta di porre fine al conflitto con una soluzione *giusta*. Le istanze percepite come non negoziabili variano di molto tra società distinte e, all'interno di esse, a seconda del codice morale di riferimento degli individui che le compongono. Ciò nonostante, vi sono circostanze in cui la scelta appare, pressoché a tutti, genuinamente tragica: quando l'interesse o il valore da proteggere sia lo stesso, e la decisione, imposta dalla scarsità di risorse, riguardi chi privilegiare tra i suoi potenziali titolari. In tali casi, non vi è alcun conflitto tra principi distinti, tra i quali individuare un vincitore, e il dilemma si trasferisce su quale soggetto, o gruppo di soggetti, vedrà garantito un bene che la società non ha dubbi a considerare fondamentale e non negoziabile. La scelta appare odiosa, e ogni soluzione sempre, almeno in parte,*ingiusta*, perché la situazione chiama in causa il principio dell'eguaglianza formale di tutti gli uomini, un valore radicato nella tradizione giuridica occidentale e democratica almeno a partire dalla rivoluzione francese. La realtà, però, impone la tragica necessità di definire i criteri in base a cui risolvere l'enigma, e di applicarli, poi, al caso concreto.

Guido Calabresi e Philip Bobbit, in un fondamentale e già menzionato libro sul tema del 1978, *Tragic Choices*, mettono in evidenza che ogni dilemma tragico sottende, in realtà, almeno due scelte[944]. La prima – da essi definita *first-order choice* - è la scelta allocativa a monte, che definisce gli indirizzi generali di una società, facendo sì che alcune prerogative, beni e servizi saranno

during clinically relevant conditions; Users are provided clear, essential information; Deployed models are monitored for performance and re-training risks are managed.

[943] Cfr. ad esempio "*dilemma*"in *Vocabolario Treccani*, https://www.treccani.it/vocabolario/dilemma/ (8 settembre 2022).

[944] Cfr. ancora G. CALABRESI, P. BOBBITT, *Tragic Choices*, New York-Londra, 1978.

disponibili in una determinata quantità. La *second-order choice*, invece, è la decisione comunemente percepita come tragica, consistente nel definire chi avrà accesso a tali risorse, qualora si verificassero condizioni di scarsità[945]. Riprendendo uno degli esempi dei due autori, è una *first-order choice*, ad esempio, l'implementazione di una politica di controllo delle nascite, che miri, attraverso misure di vario genere, a mantenere un determinato numero medio di figli per coppia. Sono *second-order choice*, invece, le decisioni atomistiche che influenzano le possibilità procreative di individui determinati, sottoforma di aperte proibizioni, o incentivi e ostacoli di ordine economico e burocratico[946]. Uno dei contributi essenziali dell'opera di Calabresi e Bobbit è chiarire che anche le *first-order choice* sono scelte tragiche, nonostante, come già rilevato, il loro contenuto etico – o la loro stessa esistenza – passi spesso in secondo piano. Esse, infatti, definiscono i limiti di risorse che sulla carta vorremmo disponibili in misura illimitata, bilanciando valori considerati incommensurabili: porre un limite al numero desiderabile di nuove nascite, ad esempio, significa stabilire i confini del diritto di ogni persona ad autodeterminarsi ed essere genitore, per ragioni di pianificazione economica o per la scarsità di risorse da destinare al welfare state.

Le opzioni ideologiche e gli orizzonti morali prevalenti giocano, ovviamente, un ruolo ineliminabile in tali valutazioni. Essi, inoltre, svolgono una funzione essenziale anche nella definizione dei criteri con cui le società cercano di razionalizzare le *second-order choice*. Calabresi e Bobbit menzionano quattro possibili approcci, nella pratica presenti sempre in combinazioni complesse: la subordinazione delle risorse scarse a logiche di mercato; la definizione di criteri chiari e pubblici, accompagnata dall'*accountability* delle autorità decidenti; il ricorso al caso, nella forma di un'estrazione, o dell'applicazione meccanica di politiche di *first-come first-served*; la "scelta di non scegliere" che porta allo sviluppo e alla continua evoluzione di prassi e convenzioni spesso incerte[947]. Le scelte tragiche sui singoli, dunque, raramente vengono demandate a un responsabile senza la formulazione, almeno implicita, di criteri-guida da parte di poteri pubblici o privati, spesso gli stessi responsabili della *first-order-choice*. D'altro canto, deve considerarsi che il ruolo del decisore, o del gruppo di decisori, incaricato della *second-order choice* è ineliminabile, al pari del contenuto etico di quest'ultima. Infatti, per quanto la decisione possa apparire vincolata da criteri predefiniti, la necessità di calare tali criteri nel caso concreto permarrà sempre. Gli studiosi dell'esegesi giuridica, d'altronde, hanno da tempo fatto luce su quanto i concetti di intepretare e

[945] Gli Autori distinguono i due concetti fin dall'introduzione del libro: «Tragic choices show two kinds of moving progressions. First, there is society's oscillation between the two sorts of decisions it must make about the scarce good. It must decide how much of it will be produced, within the limits set by natural scarcity, and also who shall get what is made. In this book the former decision is called a first-order determination and the latter a second-order determination or decision», G. CALABRESI, P. BOBBITT, *Tragic choices cit.*, p. 14.
[946] *Ibidem*.
[947] *Id.*, p. 31 ss.

decidere si sovrappongano[948]. Il dilemma tragico, quindi, non può essere occultato dietro la parvenza di un meccanicismo inesistente, il quale, peraltro, non farebbe che spacciare per neutre alcune delle scelte più complesse cui è chiamata ogni società e non sarebbe, dunque, da accogliere con favore.

Svolte queste premesse, è agevole intuire perché nell'ambito medico le scelte tragiche siano particolarmente frequenti[949]. Gli investimenti in ambito sanitario, pubblici e privati, implicano la decisione, a monte, sulla quantità di risorse terapeutiche a disposizione. Qualora tali risorse si rivelassero insufficienti, sarà necessario stabilire l'ordine di priorità del loro utilizzo, posticipando o negando ad alcuni l'accesso a determinati trattamenti medici. La scarsità di risorse può essere causata da molti fattori: investimenti consapevolmente ristretti, disponibilità oggettivamente limitata di certi strumenti, errori nella previsione di determinati bisogni sanitari, emergenze difficilmente preventivabili al momento della *first-order choice*. Il risultato, in ogni caso, è il medesimo: scelte in qualche misura tragiche – anche la compilazione di una lista d'attesa per un controllo clinico, in fondo, lo è – fanno parte della quotidianità dei sistemi sanitari e dei professionisti che vi operano.

Da questo punto di vista, l'impatto psicologico e morale della scelta tragica sull'individuo chiamato a compierla – il titolare della *second-order choice* – è stato investigato nel profondo. Essere incaricati di decisioni di tal genere è, prima di tutto, faticoso: risolvere il dilemma porta a consumare tempo ed energie che potrebbero essere destinati altrove, in primo luogo, nel caso del personale sanitario, all'attività di cura. Inoltre, il costo in termini umani, a breve e lungo termine, è evidente: l'opzione scelta è sempre percepita, almeno in parte, come moralmente sbagliata dal decidente, poiché non vi è una chiara gerarchia dei valori in gioco a cui appellarsi. Ciò genera, stress, sensi di colpa e stati depressivi[950]. La necessità di razionalizzare e limitare l'impatto di questi effetti porta allo sviluppo di strategie di gestione delle scelte tragiche alternative alla piena assunzione di responsabilità da parte del decidente[951]. La letteratura psicologica e filosofica in materia ha individuato, in primo luogo, fenomeni di c.d. *illusionism*: argomenti di varia natura – in

[948] Cfr. ad esempio, da prospettive diverse e d'ampio respiro, R. GUASTINI, *Interpretare e argomentare*, in A. CICU, F. MESSINEO, L. MENGONI, *Trattato di diritto commerciale*, Milano, XIV, 2021; T. BUSTAMANTE, C. DAHLMAN, *Argument types and fallacies in legal argumentation*, Cham, 2015; G. TUZET, *Dover decidere. Diritto, incertezza, ragionamento*, Roma, 2010.

[949] Sul tema cfr. ad es. M. R. HUNT; C. SINDING; L. SCHWARTZ, *Tragic choices in humanitarian health work*, in *The Journal of clinical ethics*, 23, 4, 2012; F. A. CARNEVALE, *Moral distress in the ICU: it's time to do something about it*, in *Minerva anestesiologica*, 86, 4, 2020; N. MESSER, *Healthcare Resource Allocation and the 'Recovery of VirtuÈ*, in *Studies in Christian Ethics*, 18, 1, 2005.

[950] Cfr. *ex multis* P.B. WHITEHEAD, R.K. HERBERTSON, A.B. HAMRIC, E.G. EPSTEIN, J.M. FISHER, *Moral distress among healthcare professionals: report of an institution-wide survey*, in *Journal of Nursing Scholarship*, 47, 2015, p. 117–125; F. A. CARNEVALE, *Moral distress in the ICU: it's time to do something about it cit.*

[951] Sul punto, cfr. in particolare J. DANAHER, *Tragic Choices and the Virtue of Techno-Responsibility Gaps*, in *Philosophy & Technology*, 35, 2, 2022.

ogni caso, a un attento esame, sempre fallaci - utilizzati per smentire la tragicità della scelta, concependola come necessitata, meramente tecnica, o, come in parte visto in precedenza, l'ultimo anello di una catena di atti dovuti[952]. In secondo luogo, di fronte alle scelte tragiche risulta particolarmente comune la tentazione, ove possibile, di delegare a terzi, ritenuti più esperti e competenti, al fine di liberarsi del peso della decisione o, almeno, condividerlo con altri[953]. La delega, tuttavia, non risolve il problema: la tragicità della scelta è semplicemente trasferita, immutata, in capo al delegato e, in ogni caso, non esiste sempre qualcuno di più preparato a cui rivolgersi. Questo è comune, in particolare, in ambito sanitario, in cui il soggetto incaricato della scelta tragica, tipicamente, è l'individuo più idoneo alla decisione, perché conosce meglio di chiunque altro le caratteristiche del trattamento e dei pazienti coinvolti.

3.4. L'ipotesi di utilizzare l'intelligenza artificiale nelle scelte tragiche in ambito sanitario e il ruolo dell'emergenza connessa alla pandemia di Covid-19

Nei sistemi sanitari avanzati, lo smistamento di pazienti che necessitano di trattamenti vitali immediati – intuitivo archetipo di ogni scelta tragica – è relegato a scenari estremamente rari, come emergenze improvvise, catastrofi naturali e contesti bellici[954]. La disponibilità di presidi terapeutici d'urgenza è, normalmente, sufficiente a fronteggiare la domanda, e il personale sanitario non si trova di fronte alla scelta di decidere a che pazienti somministrare un determinato trattamento e quali, invece, sacrificare. Scelte che implicano potenziali risvolti tragici, comunque, fanno parte dell'attività di ogni struttura sanitaria: l'esempio già citato della compilazione di liste d'attesa è particolarmente calzante. Inoltre, anche la gestione efficiente di trattamenti vitali – in primo luogo, le tempistiche d'accesso e di permanenza nei reparti di terapia intensiva –avviene in base a delicate valutazioni, generalmente di carattere predittivo. Ne è un perfetto esempio il SOFA score, acronimo di *sequential organ failure assessment score*, un'elaborazione numerica della probabilità di insufficienze d'organo dei pazienti in terapia intensiva, basata sulla combinazione di punteggi distinti, assegnati rispettivamente agli apparati respiratorio, cardiocircolatorio, epatico, neurologico, escretore e alla capacità di coagulazione[955]. Questo genere di valutazioni permette, ad esempio, di

[952] Cfr. ancora J. DANAHER, *Tragic Choices and the Virtue of Techno-Responsibility Gaps cit.*; J. COOPER, *Cognitive Dissonance: Where We've Been and Where We're Going*, in *International Review of Social Psychology*, 32, 1, 2019. Più in generale v. S. SMILANSKY, *Free Will and Illusion*, Oxford, 2000.

[953] J. RAZ,*Law, morality and authority*, in*The Monist*, 68, 3, 1985, p. 295–324 si spinge fino ad individuare nella necessità di liberare il singolo individuo dal penso di determinate scelte uno dei fondamenti principali dell'autorità delle istituzioni formali. Cfr. inoltre M. STEFFEL, E.F. WILLIAMS, *Delegating decisions: recruiting others to make choices we might regret*, in *Journal of Consumer Research*, 44, 5, 2018, p. 1015-1032.

[954] Oltre ai numerosissimi studi sul tema, alcuni dei quali già citate, un riferimento d'altro genere, ma imprescindibile, in materia di tragicità del *triage* medico è la sofferta testimonianza di G. STRADA, *Pappagalli verdi. Cronache di un chirurgo di guerra*, Milano, 1999.

[955] Cfr. ad esempio M. SINGER ET AL., *The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)*, in *Journal of American Medical Association*, 315, 8, 2016, p. 801-810; F.L. FERREIRA, D.P. BOTA, A.

identificare precocemente potenziali emergenze, ipotizzare l'esito di determinate terapie, calcolare la durata stimata della degenza dei pazienti. In tal modo, le risorse disponibili sono utilizzate in modo più efficiente – evitandone, normalmente, la saturazione - e il lavoro del personale sanitario risulta più efficace.

L'utilità di elaborazioni predittive per l'attività ospedaliera su larga scala, e in particolare per la gestione di reparti di terapia intensiva e altri trattamenti d'urgenza, ha fatto ipotizzare che l'intelligenza artificiale basata sull'analisi dei dati potesse dimostrarsi molto efficace nell'ambito. In particolare, come già accennato nei paragrafi precedenti, le possibilità dischiuse dall'apprendimento automatico hanno portato allo sviluppo di modelli in grado di formulare previsioni del decorso clinico e dei bisogni a breve e medio termine dei pazienti più accurate di quelle ottenibili con metodi come il SOFA score. Infatti, le capacità di analisi di una mole di dati sempre crescente, garantite dalle reti neurali, permettono di costruire sistemi in grado di elaborare più variabili, e tenere conto di più precedenti, rispetto alle tecniche di valutazione numerica, anche complesse, già in uso in ambito sanitario, che non prevedono l'utilizzo di intelligenza artificiale[956]. Strumenti di questo genere sono stati l'oggetto di svariati articoli scientifici nella seconda metà degli anni '10 del 2000, in coincidenza con il generale picco d'interesse dimostrato da industria e ricerca per le tecnologie basate sul *deep learning*[957]. Sul loro effettivo impiego in strutture sanitarie, invece, non è agevole reperire informazioni[958]. Fin da subito, alcune voci hanno evidenziato le delicatissime questione etiche poste da tali applicazioni dell'intelligenza artificiale, poiché la possibilità di

---

BROSS, C. MÉLOT, J.L. VINCENT, *Serial Evaluation of the SOFA Score to Predict Outcome in Critically Ill Patients*, in *Journal of American Medical Association*, 286, 14, 2001; J.L. VINCENT ET AL., *The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure*, in *Intensive Care Medicine*, 22, 7, 1996.

[956] Sul punto, cfr. approfonditamente *supra*, p. 140 ss. p. 281 ss.

[957] Cfr. ad esempio A. RAJKOMAR ET AL., *Scalable and accurate deep learning with electronic health records*, in *Nature Digital Medicine*, 1, 1, 2018; P. NGUYEN, T. TRAN, S. VENKATESH, *Deep Learning to Attend to Risk in ICU*, 2017, https://arxiv.org/abs/1707.05010 (9 settembre 2022); R. S. ANAND ET AL., *Predicting Mortality in Diabetic ICU Patients Using Machine Learning and Severity Indices*, in *Proceeding from AMIA Summits on Translational Science Proceedings*, 2018, p. 310-319; S.Y. KIM ET AL., *A deep learning model for real-time mortality prediction in critically ill children*, in *Critical Care*, 23, 1, 2019; E. MACIAS, G. BOQUET, J. SERRANO, J.L. VICARIO, J. IBEAS, A .MORELL, *Novel Imputing Method for the Early Prediction of Sepsis in ICU Using Deep Learning Techniques*, Computing in Cardiology, 2019, https://bit.ly/3ynxIq3 (9 settembre 2022); A. BHATTACHARYYA, S. SHEIKHALISHAHI, S. DUGAR, S. KRISHNAN, A. DUGGAL, V. OSMANI, *Predicting delirium risk for the following 24 hours in critically ill patients using deep learning*, in *Journal of Critical Care Medicine*, 48, 1, 2020, p. 182, 10.1097/01.ccm.0000619952.70488.fb; B. MAMANDIPOOR; M. MAJD; M. MOZ, V. OSMANI, *Blood Lactate Concentration Prediction in Critical Care*, in *Digital Personalized Health and Medicine*, 2020, doi:10.3233/SHTI200125; B. MAMANDIPOOR, F. FRUTOS-VIVAR, O. PEÑUELAS, R. REZAR ET AL., *Machine learning predicts mortality based on analysis of ventilation parameters of critically ill patients: multi-centre validation*, in *BMC Medical Informatics and Decision Making Journal*, 2021, doi: 10.1186/s12911-021-01506-w; B. MAMANDIPOOR, W. YEUNG, L. AGHA-MIR-SALIM ET AL.,*Prediction of blood lactate values in critically ill patients: a retrospective multi-center cohort study*, in *Journal of Clinical Monitoring and Computing*, 36, 2022, p. 1087–1097, https://doi.org/10.1007/s10877-021-00739-4.

[958] Gli avanzamenti della ricerca, infatti, hanno attirato l'attenzione della stampa, ma non risultano notizie di effettivi utilizzi ospedalieri, né comunicazioni in proposito di strutture sanitarie. Cfr. M. MAGISTRONI, *Intelligenza artificiale: un algoritmo predice quanto tempo resta ai malati terminali*, Wired, 22 gennaio 2018, https://bit.ly/3rIbbRh (9 settembre 2022); R. MANTOVANI, *Ricordati che devi morire? Ci pensa Google*, Focus, 25 giugno 2018; *L'Intelligenza Artificiale predice morti premature meglio dell'uomo*, Sky TG24, 28 marzo 2019, https://bit.ly/3EBXmLE (9 settembre 2022).

elaborare predizioni sempre più approfondite sul decorso clinico del paziente implica, in caso di prognosi infausta, anche quella di identificare con crescente precisione il momento e la probabilità della morte di quest'ultimo[959]. Lo scenario è stato, anche da questo punto di vista, rivoluzionato dalla pandemia di Covid-19.

Il propagarsi dell'epidemia in tutto il mondo, nei primi mesi del 2020, ha portato, in diversi sistemi sanitari, a una situazione di estrema scarsità. In particolare, la polmonite bilaterale che colpisce, nei casi più gravi, i soggetti infettati dal virus Sars-Cov-2 ha portato all'aumento esponenziale dei pazienti che necessitavano l'utilizzo di respiratori per la ventilazione meccanica o il ricovero in reparti di terapia intensiva. I supporti medici necessari per la ventilazione hanno rapidamente cominciato a scarseggiare e le terapie intensive a saturarsi, mettendo a rischio la possibilità di accedervi dei malati di altre patologie. L'emergenza che ne è derivata ha obbligato il personale sanitario alle scelte tragiche radicali che, al principio di questo paragrafo, abbiamo definito tendenzialmente molto rare: il *triage* tra chi avrà accesso a un determinato trattamento salvavita e chi, invece, non potrà usufruirne[960]. Nonostante non esistano dati precisi sul numero di pazienti coinvolti, la situazione ha riguardato la maggioranza dei sistemi sanitari europei, del continente americano e diversi paesi asiatici, mentre altre aree del mondo, colpite dalla pandemia solo più tardi o caratterizzate da una popolazione particolarmente giovane, hanno dimostrato una migliore tenuta[961]. Gli interrogativi etici che hanno accompagnato questa situazione d'emergenza sono intuibili, e i criteri utilizzati per effettuare la selezione – non sempre esplicitati – sono stati sottoposti ad approfondite discussioni. Com'è noto, l'Italia è stata tra i paesi in cui l'epidemia si è diffusa prima e più velocemente, e la necessità di razionare le limitate risorse sanitarie a disposizione è stata particolarmente severa. Un documento predisposto, il 6 marzo 2020, dalla *Società Italiana di Anestesia, Analgesia, Rianimazione e Terapia Intensiva*(SIAARTI), contenente alcune linee guida per lo svolgimento in tempi rapidi del *triage* d'urgenza, è stato discusso anche in

[959]Finalità, del resto, che alcuni tra gli studi citati dichiarano esplicitamente, cfr. ad es. A. RAJKOMAR ET AL., *Scalable and accurate deep learning with electronic health records cit.*; R. S. ANAND ET AL., *Predicting Mortality in Diabetic ICU Patients cit.*; S.Y. KIM ET AL., *A deep learning model for real-time mortality prediction cit.* Per degli esempi del recente dibattito etico sul tema cfr. J.A. SHAW, N. SETHI, B.L. BLOCK, *Five things every clinician should know about AI ethics in intensive care*, in *Intensive Care Medicine*, 47, 2, 2021, p. 157-159; M. BEIL; I. PROFT; D. VAN HEERDEN; S. SVIRI; P. V. VAN HEERDEN, *Ethical considerations about artificial intelligence for prognostication in intensive care*, in *Intensive Care Medicine Experimental*, 7, 1, 2019.

[960] Cfr. da vari punti di vista K. ORFALI, *What Triage Issues Reveal: Ethics in the COVID-19 Pandemic in Italy and France*, in *Journal of Bioethical Inquiry*, 17, 4, 2020; K. ORFALI, *Getting to the Truth: Ethics, Trust, and Triage in the United States versus Europe during the Covid-19 Pandemic*, in *Hastings Center Report*, 51, 1, 2021; P. ERIKA ET AL., *Triage decision-making at the time of COVID-19 infection: the Piacenza strategy*, in *Internal and Emergency Medicine*, vol. 15, 5, 2020; S. SPINSANTI, *Il Covid-19 ci ha costretto a confrontarci col triage*, Corriere della Sera, 21 marzo 2022;

[961]Per una panoramica sulla storia e gli sviluppi della pandemia di Sars-Cov-2, a quasi tre anni dal suo inizio, cfr. *Cose che noi umani. La pandemia che ha sconvolto le nostre vite e resterà per sempre nell'immaginario comune*. *Una cronistoria degli eventi che non avremmo mai potuto immaginare*, Lab24 – Il Sole 24 Ore, https://bit.ly/3TcWDF3 (9 ottobre 2022); B. WALSH, *Covid-19: The history of pandemics*, in *BBC future*, 26 marzo 2020, https://bbc.in/3VdS1QL (9 ottobre 2022).

campo internazionale, e riassume in modo efficace il dibattito che ha caratterizzato quei giorni[962]. Una frase, in particolare, è stata messa pesantemente in discussione, poiché non prevedeva eccezioni e sembrava introdurre automatismi che sminuivano il valore intrinseco dell'essere umano: «può rendersi necessario porre un limite di età all'ingresso in terapia intensiva»[963]. Per quanto non siano mancate voci, anche prestigiose, a difesa del criterio, giudicato pragmatico e adatto ad essere applicato in condizioni di estremo stress ed emergenza, esso è stato prima criticato dal Comitato Nazionale per la Bioetica[964], poi corretto dalla stessa SIAARTI, in un documento dell'anno successivo[965]. In ambo i casi, è stata riconosciuta la necessità di una valutazione globale delle prospettive di guarigione, delle condizioni sanitarie di partenza del soggetto e dell'ipotizzabile quantità e qualità degli anni di vita a cui avrà accesso in caso di esito positivo delle terapie, e affermato che in nessun caso l'età può ergersi a unico criterio di selezione.

La necessità di operare la selezione elaborando, in tempi molto rapidi, una pluralità di variabili ha generato un crescente interesse per le possibilità connesse all'applicazione, a riguardo, dell'intelligenza artificiale. Fin dall'inizio dell'emergenza, la ricerca ha rivolto i propri sforzi verso lo sviluppo di sistemi di apprendimento automatico analoghi a quelli già visti, "allenati" con dati relativi ai pazienti del nuovo virus. Sono stati sviluppati, in particolare, sistemi di automatizzazione della diagnosi della malattia, basati sull'*image recognition* di scansioni polmonari o su altri elementi[966]; strumenti predittivi del decorso clinico del paziente e, in particolare, della necessità di

[962] SIAARTI, *Raccomandazioni di etica clinica per l'ammissione a trattamenti intensivi e per la loro sospensione in condizioni eccezionali di squilibrio tra necessità e risorse disponibili*, 6 marzo 2020, https://bit.ly/3fW0SX9 (9settembre 2022). Per alcuni commenti, da varie prospettive, cfr. M. G. BERNARDINI, *Una questione di interpretazione? Note critiche su Raccomandazioni SIAARTI, discriminazione in base all'età ed emergenza sanitaria*, in *BioLaw Journal - Rivista di BioDiritto*, 3, 2020; A. RIMEDIO, *Criteri di priorità per l'allocazione di risorse sanitarie scarse nel corso della pandemia da CoViD-19*, in *BioLaw Journal - Rivista di BioDiritto*, 1, 2021; L. D'AVACK, *CoViD-19: criteri etici*, in *BioLaw Journal - Rivista di BioDiritto*, 1S, 2020; G. RAZZANO,*Riflessioni a margine delle raccomandazioni SIAARTI per l'emergenza Covid-19, fra triage, possibili discriminazioni e vecchie DAT: verso una rinnovata sensibilità per il diritto alla vita?*, in *Rivista AIC*,3, 2020, p. 107-129; C. DELLA GIUSTINA, *Il problema della vulnerabilità nelle Raccomandazioni SIAARTI e nelle linee guida SIAARTI-SIMLA*, in *Stato, Chiese e pluralismo confessionale*, 2021, https://bit.ly/3Vf3gZ3 (9 settembre 2022); K. ORFALI, *What Triage Issues Reveal: Ethics in the COVID-19 Pandemic in Italy and France cit.*

[963] SIAARTI, *Raccomandazioni di etica clinica per l'ammissione a trattamenti intensivi cit.*, p. 5.

[964] Cfr. COMITATO NAZIONALE PER LA BIOETICA, *Covid-19: la decisione clinica in condizioni di carenza di risorse e il criterio del "triage in emergenza pandemica"*, 8 aprile 2020, in cui l'età è considerata uno tra i molti elementi per la valutazione clinica del paziente, e si esclude che possa essere l'unico o il principale di essi. Per una difesa dell'impostazione adottata dalle Linee Guide SIAARTI, si veda invece l'opinione di minoranza, pubblicata unitamente al parere del CNB, del prof. Maurizio Mori.

[965] SIAARTI-SMILA,*Decisioni per le cure intensive in caso di sproporzione tra necessità assistenziali e risorse disponibile in corso di pandemia di Covid-19*, 13 gennaio 2021, https://bit.ly/3ROEkok (10 settembre 2022). Il documento, tra le altre cose, contiene un'affermazione molto netta in merito all'età: «l'età deve essere considerata nel contesto della valutazione globale della persona malata e non sulla base cut-off predefiniti».

[966] Cfr. ad es. F. ZHANG, *Application of machine learning in CT images and X-rays of COVID-19 pneumonia*, in*Medicine*,100, 36, 2021,https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8428739/ (10 settembre 2022); X. FAN, X. FENG, Y. DONG, H. HOU, *COVID-19 CT image recognition algorithm based on transformer and CNN*, in*Displays*, 72, 2022; N. SUBRAMANIAN, O. ELHARROUSS, S. AL-MAADEED, M. CHOWDHURY,*A review of deep learning-based detection methods for COVID-19*, in*Computers in Biology and Medicine*, 143, 2022, 105233; G. JAIN, D. MITTAL, D. THAKUR, M.K. MITTAL, *A deep learning approach to detect Covid-19 coronavirus with X-Ray images*, in*Biocybernetics*

ventilazione artificiale o dell'accesso ai reparti di terapia intensiva[967]; sistemi, strettamente connessi a quelli appena menzionati, che, in via esclusiva o accanto a valutazioni sull'evoluzione delle sue condizioni, formulavano previsioni attinenti alle probabilità e al momento della morte del paziente[968].

Nonostante alcuni articoli della stampa generalista, nel corso dell'emergenza, abbiano ipotizzato la circostanza la circostanza, non è agevole determinare, anche in questo caso, se tali algoritmi abbiano avuto effettivo utilizzo ospedaliero o potranno averlo in futuro, qualora si presentassero nuove crisi[969]. Non si può trascurare, ad ogni modo, che la loro potenzialeutilità – al pari, del resto, degli strumenti analoghi non attinenti all'emergenza pandemica–non sarebbe limitata alla gestione del *triage*: le loro previsioni, infatti, permetterebbero una pianificazione più accuratadelle risorse nel medio periodo, e incrementerebbero l'efficacia globale dell'intervento sanitario, portando all'identificazione precoce dei pazienti a rischio di complicazioni. D'altro canto, il coinvolgimento di tecnologie avanzate in scelte tragiche come quelle in esame pone interrogativi all'etica e al diritto che non possono trascurarsi, anche in ragione di alcune caratteristiche tecniche delle reti neurali, come la loro scarsa interpretabilità. Il rischio di non rispettare la dignità umana e i diritti dei soggetti coinvolti, infatti, appare particolarmente alto, anche in ragione della situazione di emergenza in cui è presa la decisione. Proprio all'analisi dei risvolti etico-giuridici della combinazione tra scelte tragiche e sistemi intelligenti, prima di tutto dal punto di vista dei diritti fondamentali, sarà dedicato il prossimo paragrafo.

---

and Biomedical Engineering, 40, 4, 2020. Per alcuni limiti di tali sistemi, cfr. G. MAGUOLO, L. NANNI, *A critic evaluation of methods for COVID-19 automatic detection from X-ray images*, in*Information Fusion*, 76, 2021, p. 1-7.

[967]S. P. SHASHIKUMAR ET AL., *Development and Prospective Validation of a Deep Learning Algorithm for Predicting Need for Mechanical Ventilation*, in*Chest*, 159, 6, 2021; J. P. COHEN ET AL., *Predicting COVID-19 Pneumonia Severity on Chest X-ray With Deep Learning*, in*Cureus*,12, 7, 2020; N. LASSAU ET AL., *Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID-19 patients*, in*Nature Communications*, 12, 1, 2021, https://www.nature.com/articles/s41467-020-20657-4 (10 settembre 2022).

[968]Si vedano ad esempio C. JUNG ET AL., *Disease-Course Adapting Machine Learning Prognostication Models in Elderly Patients Critically Ill With COVID-19: Multicenter Cohort Study With External Validation*, in*JMIR Medical Informatics*, 10, 3, 2022, https://medinform.jmir.org/2022/3/e32949 (10 settembre 2022); X. LI ET AL., *Deep learning prediction of likelihood of ICU admission and mortality in COVID-19 patients using clinical variables*, in*PeerJ*, 8, 2020, https://peerj.com/articles/10337 (10 settembre 2022); J. CHENG ET AL., *COVID-19 mortality prediction in the intensive care unit with deep learning based on longitudinal chest X-rays and clinical data*, in*European Radiology*, 32, 7, 2022; L. GOURDEAU ET AL., *Deep learning of chest X-rays can predict mechanical ventilation outcome in ICU-admitted COVID-19 patients*, in *Scientific Reports*, 12, 1, 2022, https://www.nature.com/articles/s41598-022-10136-9 (10 settembre 2022); C. LIEW, J. QUAH, H. L. GOH, N. VENKATARAMAN, *A chest radiography-based artificial intelligence deep-learning model to predict severe Covid-19 patient outcomes: the CAPE (Covid-19 AI Predictive Engine) model*, in *MedRxiv*, https://doi.org/10.1101/2020.05.25.20113084 (10 settembre 2022).

[969] Cfr. ad esempio K. HAO, *Doctors are using AI to triage covid-19 patients. The tools may be here to stay*, in *MIT Technology Review*, 23 aprile 2020; L. CUPPINI, *Covid, il rischio di morte dei pazienti calcolato da un algoritmo*, Corriere della Sera, 5 settembre 2020. Non vi è traccia, però, di testimonianze dell'utilizzo di tali strumenti in ambito ospedaliero in articoli sottoposti a *peer review* o in documenti ufficiali.

3.5. Il quadro normativo applicabile all'eventuale utilizzo dell'intelligenza artificiale nelle scelte tragiche, le prospettive aperte dalla Proposta di Regolamento in materia di intelligenza artificiale e il ruolo centrale dei diritti fondamentali, vecchi e nuovi

L'impiego di tecnologie avanzate in scelte tragiche, e più in generale nella valutazione diagnostica e terapeutica, pone il problema, in primo luogo, della consapevolezza del paziente. È circostanza nota che il rapporto tra curante e curato è stato, per secoli, fortemente paternalistico: l'individuo che necessitasse assistenza sanitaria si affidava completamente al medico, spesso comprendendo molto poco delle proprie condizioni e delle terapie scelte per affrontarle. La situazione è gradualmente evoluta anche in seguito a ripetuti scandali, in cui trattamenti sanitari svolti, per diverse finalità, senza il consenso dei soggetti interessati hanno causato gravi pregiudizi a singoli o interi gruppi sociali, spesso in condizioni di marginalità[970]. Si è imposta una concezione dell'arte medica che assegna al paziente il ruolo centrale, e la necessità di un consenso effettivo, libero e informato a ogni misura sanitaria che lo riguardi si è imposta prima come principio bioetico, poi come norma giuridica, anche nell'ordinamento italiano[971]. È lecito chiedersi, allora, se le informazioni da riferire al paziente debbano riguardare anche l'utilizzo di tecnologie a supporto della decisione del curante: un requisito riconducibile, nella sistematica di nuovi diritti proposta in questo lavoro, al menzionato diritto alla *disclosure* dell'interazione con sistemi intelligenti.

La completezza delle norme in materia di consenso ai trattamenti sanitari porta a ritenere che il paziente debba essere informato anche dell'utilizzo di sistemi di intelligenza artificiale. Per quanto riguarda l'ordinamento italiano, ad esempio, l'art. 1 c. 3 della L. 217 del 2019[972] - volta, com'è

[970] Oltre agli scandali che hanno caratterizzato i regimi instuaratisi in Europa nella prima metà del XX secolo, in primo luogo la Germania nazista, nel secondo dopoguerra hanno assunto particolare notorietà gli esperimenti condotti dallo United States Public Health Service a Tuskegee, in Alabama, in cui a cittadini afroamericani è stato precluso l'accesso ai trattamenti antisifilide anche dopo che la penicillina aveva dimostrato la sua efficacia per la cura della malattia, e le campagne di sterilizzazione obbligatoria condotte in Svezia – ma si rinvengono esempi analoghi anche nella storia recente di altri paesi europei - per motivi medici, eugenetici e sociali fino al 1976, cfr. J. HELLER, *Syphilis victims in U.S. study went untreated for 40 years*, New York Times, 26 luglio 1972; R.V. KATZ, S.S. KEGELES, N.R. KRESSIN, B. LEE GREEN, M.Q. WANG, S.A. JAMES, S.L. RUSSELL, C. CLAUDIO, *The Tuskegee Legacy Project: willingness of minorities to participate in biomedical research*, in *Journal of healthcare for the poor and underserved*, 17, 4, 2006, p. 698-715; V. LANZA, *Svezia, sterilizzate a forza*, la Repubblica, 25 agosto 1997; M. CONGIU, *Repubblica Ceca, legge su risarcimento donne Rom sottoposte a sterilizzazione forzata*, il Manifesto, 1 luglio 2021.

[971] Il consenso del soggetto compare tra i principi essenziali enunciati da due dei principali documenti elaborati in campo bioetico dopo la seconda guerra mondiale, il Codice di Norimberga del 1947, frutto del c.d. "processo ai dottori", il giudizio che vide sul banco degli imputati, nel processo di Norimberga, i medici colpevoli di sperimentazioni e torture condotte nei lager nazisti, e il c.d. *Belmont Report* del 1979, *corpus* di principi bioetici elaborato dallo United States Department of Health and Human Services a seguito dell'appena menzionato scandalo di Tuskegee (cfr. nota precedente).

[972] Legge n. 219 del 22 dicembre 2017, *Norme in materia di consenso informato e di disposizioni anticipate di trattamento*. I primi due commi del citato articolo 1 danno un quadro esaustivo delle finalità che ispirano il provvedimento normativo: «La presente legge, nel rispetto dei principi di cui agli articoli 2, 13 e 32 della Costituzione e degli articoli 1, 2 e 3 della Carta dei diritti fondamentali dell'Unione europea, tutela il diritto alla vita, alla salute, alla dignita' e all'autodeterminazione della persona e stabilisce che nessun trattamento sanitario puo' essere iniziato o proseguito se privo del consenso libero e informato della persona interessata, tranne che nei casi espressamente previsti dalla legge. 2. È promossa e valorizzata la relazione di cura e di fiducia tra paziente e

noto, proprio a valorizzare la volontà del paziente e la relazione di cura –dispone che «ogni persona ha il diritto di conoscere le proprie condizioni di salute e di essere informata in modo completo, aggiornato e a lei comprensibile riguardo alla diagnosi, alla prognosi, ai benefici e ai rischi degli accertamenti diagnostici e dei trattamenti sanitari indicati, nonchè riguardo alle possibili alternative e alle conseguenze dell'eventuale rifiuto del trattamento sanitario e dell'accertamento diagnostico o della rinuncia ai medesimi». Alla lettera, la norma non contiene alcun riferimento alle tecnologie coinvolte nel trattamento, al pari di quelle ad essa accostabili, di fonte straniera o sovranazionale[973]. L'unica eccezione di rilievo, come si dirà, è rappresentata ancora una volta dalla Francia, che ha di recente emanato alcune disposizioni specifiche riguardanti l'utilizzo di algoritmi nella decisione medica, comprendenti specifici doveri di informazione del paziente[974]. Tuttavia, l'intensità e il dettaglio degli obblighi informativi previsti dalle norme in materia di consenso sanitario, in primo luogo quelle italiane appena citate, rende difficilmente immaginabile che essi possano essere soddisfatti omettendo la presenza di strumenti intelligenti utilizzati per la diagnosi, la decisione del trattamento più adeguato o la valutazione delle probabilità del suo successo.

Sullo scenario europeo, un elemento ulteriore da cui sembra possibile ricavare l'esistenza di oneri d'informazione di questo tipo viene dalle norme, già analizzate più volte, in materia di decisione automatizzata. Le informazioni da fornire al paziente sul trattamento dei suoi dati personali – da tenere distinte, dal punto di vista giuridico, da quelle sul trattamento sanitario – dovranno riguardare, allora, anche le tecnologie intelligenti coinvolte, ai sensi degli artt. 13, 14, 15 e 22

---

medico che si basa sul consenso informato nel quale si incontrano l'autonomia decisionale del paziente e la competenza, l'autonomia professionale e la responsabilita' del medico. Contribuiscono alla relazione di cura, in base alle rispettive competenze, gli esercenti una professione sanitaria che compongono l'equipe sanitaria. In tale relazione sono coinvolti, se il paziente lo desidera, anche i suoi familiari o la parte dell'unione civile o il convivente ovvero una persona di fiducia del paziente medesimo».

[973] È il caso, ad esempio, dell'art. 4 della Legge spagnola 41/2002, *básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica*: «1. Los pacientes tienen derecho a conocer, con motivo de cualquier actuación en el ámbito de su salud, toda la información disponible sobre la misma, salvando los supuestos exceptuados por la Ley. Además, toda persona tiene derecho a que se respete su voluntad de no ser informada. La información, que como regla general se proporcionará verbalmente dejando constancia en la historia clínica, comprende, como mínimo, la finalidad y la naturaleza de cada intervención, sus riesgos y sus consecuencias. 2. La información clínica forma parte de todas las actuaciones asistenciales, será verdadera, se comunicará al paciente de forma comprensible y adecuada a sus necesidades y le ayudará a tomar decisiones de acuerdo con su propia y libre voluntad. 3. El médico responsable del paciente le garantiza el cumplimiento de su derecho a la información. Los profesionales que le atiendan durante el proceso asistencial o le apliquen una técnica o un procedimiento concreto también serán responsables de informarle», o, sul piano internazionale, dell'art. 5 della *Convenzione di Oviedo per la protezione dei Diritti dell'Uomo e della dignità dell'essere umano* nei confronti dell'applicazioni della biologia e della medicina del 4 Aprile 1997: «Un intervento nel campo della salute non può essere effettuato se non dopo che la persona interessata abbia dato consenso libero e informato. Questa persona riceve innanzitutto una informazione adeguata sullo scopo e sulla natura dell'intervento e sulle sue conseguenze e i suoi rischi. La persona interessata può, in qualsiasi momento, liberamente ritirare il proprio consenso». Sulla Convezione di Oviedo, in letteratura, cfr. ad esempio I.R. PAVONE,*La convenzione europea sulla biomedicina*, Milano, 2009.

[974] Si tratta, come si dirà (cfr. *infra*, p. 301), dell'art. L.4001-3 del *Code de la santé publique*, introdotto con la Loi n. 2021-1017 du 2 août 2021 *relative à la bioéthique*.

GDPR[975]. Rimane fermo, anche in questo caso, il principale limite di tali norme: risultano applicabili solamente quando la decisione sia totalmente delegata all'algoritmo. Un'eventualità che, per la verità, non pare particolarmente calzante per i sistemi in esame, accanto ai quali è ragionevole presumere che rimanga presente, almeno nel caso di un loro utilizzo nel breve e medio periodo, un medico in carne ed ossa (dovranno, semmai, indagarsi le genuine possibilità e capacità di questi di interferire nelle loro valutazioni). In caso di definitiva approvazione, sullo scenario europeo il quadro sarà completato dalla Proposta di Regolamento in materia di intelligenza artificiale, il cui art. 52, com'è noto, impone di comunicare l'utilizzo di un sistema intelligente destinato a interagire con le persone fisiche ogni volta che possa risultare non evidente[976]. La norma, di carattere generale, troverebbe, ovviamente, applicazione anche in ambito sanitario. La possibilità di ricondurre al suo ambito applicativo algoritmi predittivi e decisionali, però, appare piuttosto incerta[977].

Al netto di questi ultimi limiti, il diritto ad essere informati dell'utilizzo di un sistema intelligente sembra tutelato da presidi giuridici significativi, in particolare in Europa. Allo stesso tempo, però, risulta problematico immaginare strategie affinché tale comunicazione risulti efficace, e indagare il rapporto tra utilizzo dell'intelligenza artificiale e consenso al trattamento sanitario. Fornire informazioni sufficienti sulle tecnologie utilizzate, infatti, pare un compito estremamente complicato, indissolubilmente correlato alle caratteristiche socioculturali del paziente e che richiede competenze esterne all'area medica. Si tratta, del resto, di un problema destinato a presentarsi, con la diffusione delle tecnologie basate sull'intelligenza artificiale, in molti altri settori[978]. L'ambito medico pone, però, questioni ulteriori, specialmente quando si prenda in considerazione l'ipotesi di impiegare strumenti basati sull'apprendimento automatico a supporto delle scelte che abbiamo definito tragiche. Se, infatti, l'ipotesi di un'eventuale paziente che negasse il consenso a un sistema utilizzato nell'attività diagnostica – per ipotesi, una delle tecnologie avanzate basate sull'*image recognition* viste in precedenza – pare facilmente risolvibile rinunciando al suo supporto, così non è quando tali tecnologie siano applicate nella razionalizzazione di risorse sanitarie scarse. Quali sarebbero le conseguenze di tale rifiuto su un processo di selezione, magari in una situazione d'emergenza, in cui lo strumento sia usato su larga scala? Ovviamente, per il medico è sempre possibile replicare, in totale autonomia, la valutazione che avrebbe svolto col supporto dell'algoritmo (fermo quanto analizzato sui possibili effetti a lungo termine della distorsione

[975] Per il testo dei quali si rinvia *supra*, p. 183 e 163.
[976] Per il testo completo della norma cfr. *supra*, p. 183.
[977] Posto il loro utilizzo a supporto dell'attività e delle valutazioni del medico, infatti, non pare sempre pacifico che essi possano rientrare, nella prospettiva del paziente tra «i sistemi di IA destinati a interagire con le persone fisiche» di cui parla l'art. 52 della Proposta di Regolamento.
[978] Si rimanda a quanto già osservato riguardo agli oneri di informazione imposti ai funzionari pubblici francesi dalla Loi n. 2018-493 du 20 juin 2018 *relative à la protection des données personnelles*, commentata *supra*, p. 206 ss.

dell'automazione e i rischi di *deskilling*)[979]. Ma come comparare questa elaborazione con quelle – più complesse, perché basate su più variabili, e magari frutto di strumenti che si comportano come *black-box* –cui ha partecipato anche la tecnologia, relative agli altri pazienti? L'essere umano sarebbe chiamato a definire l'ordine di priorità tra misurazioni eterogenee, e non è agevole individuare i criteri per gestire questa incertezza. Non possono trascurarsi, poi, gli effetti del condizionamento psicologico del medico che il rifiuto della tecnologia potrebbe generare, o quelli delle condizioni di fretta e stress in cui egli potrebbe essere chiamato a scegliere.

In ogni caso, in materia di scelte tragiche, non sempre la situazione concreta rende ipotizzabile il consenso espresso del paziente, le cui condizioni risultano tipicamente molto gravi e impongono ritmi di lavoro serrati. Un perfetto esempio di questa circostanza è l'ipotesi, già menzionata, di dover svolgere un *triage* tra i pazienti per la scarsità di supporti sanitari di sostegno vitale, generata da una situazione di emergenza. L'eventuale utilizzo di tecnologie intelligenti in tali circostanze non solleva, ovviamente, interrogativi solo per la relazione tra paziente e curante, ma mette in discussione, in modo più basilare, l'intera struttura della decisione medica.

Pressochè tutti i sistemi in esame, infatti, si basano sull'apprendimento profondo, e i loro stati interni risultano, così, in tutto o in parte inconoscibili.L'attuale quadro normativo europeo, che riconosce una protezione solo parziale ai requisiti di spiegabilità e controllo umano sul sistema, è già stato più volte richiamato, al pari della *Directive on automated decision making canadese*, ad esso in buona parte accostabile. Anche i possibili sviluppi che potrebbero derivare dall'approvazione della Proposta di Regolamento in materia di intelligenza artificiale presentata dalla Commissione il 21 aprile 2022 sono già stati analizzati. Completa il quadro una recente iniziativa legislativa francese, cui si è già fatto cenno in materia di consenso informato, con la quale è stata introdotta, attraverso una riforma della nota *Loi de bioéthique*[980], una disciplina specifica dell'utilizzo di intelligenza artificiale basata sull'analisi dei dati a supporto di decisioni e

[979] Per il possibile deterioramente della capacità dei sanitari di prendere decisioni complesse a causa della diffusione dell'intelligenza artificiale, cfr. in particolare B.P. GREEN, *Artificial intelligence, decision-making and moral deskilling*, in *Markkulla center for applied ethics*, 15 maggio 2019, https://bit.ly/3SWwjiB (12 settembre 2022); S.J. SHAIKH, *Artificial Intelligence and resource allocation in health care: The process-outcome divide in perspectives on moral decision-making*, in *Ceur – WP*, 2884, 112, http://ceur-ws.org/Vol-2884/paper_122.pdf (12 settembre 2022). Si vedano inoltre i già citatiJ. LU, *Will Medical Technology Deskill Doctors? cit.*; J. LEVY, A. JOTKOWITZ, I. CHOWERS, *Deskilling in Ophthalmology Is the Inevitable Controllable? Cit.*; S. DE PAOLI, *Automatic-Play and Player Deskilling cit.*; E. SINAGRA, F. ROSSI, D. RAIMONDO, *Use of Artificial Intelligence in Endoscopic Training: Is Deskilling a Real Fear? cit.*

[980] Le prime leggi francesi sulla bioetica sono state emanate nel 1994, anno che ha visto la promulgazione della Loi n. 94-548 du 1° juillet 1994 *relative au traitement des données nominatives ayant pour fin la recherche dans le domaine de la santé et modifiant la loi n° 78-17 du 6 janvier 1978 relative à l'informatique et aux fichiers et aux libertés*; della Loi n 94-653 du 29 juillet 1994 *relative au respect du corps humain* e della Loi 94-654 du 29 juillet 1994 *relative au don et à l'utilisation des éléments et produits du corps humain, à l'assistance médicale à la procréation et au diagnostic prénatal*. Le leggi stesse disponevano una loro revisione entro il termine di 5 anni, poi elevati a 7. Da allora, il Parlamento ha adempiuto più volte a tale dovere di aggiornamento, con diverse *loi de bioétique*, l'ultima delle quali, la già citata Loi n. 2021-1017 du 2 août 2021 *relative à la bioéthique* (cfr. *supra*, n. 974), ha introdotto la novella del *Code de la santé publique* ora in esame.

valutazioni del personale sanitario. Tale norma, allo stato dell'arte un *unicum* nel panorama giuridico mondiale, impone netti requisiti di spiegabilità degli algoritmi, che paiono poter mettere al bando, se interpretati rigidamente, le tecnologie c.d. *black box*. Il nuovo art. L. 4001-3 del *Code de la santé publique*, introdotto con la novella, infatti, ai primi tre commi recita: «*I.-Le professionnel de santé qui décide d'utiliser, pour un acte de prévention, de diagnostic ou de soin, un dispositif médical comportant un traitement de données algorithmique dont l'apprentissage a été réalisé à partir de données massives s'assure que la personne concernée en a été informée et qu'elle est, le cas échéant, avertie de l'interprétation qui en résulte.II.-Les professionnels de santé concernés sont informés du recours à ce traitement de données. Les données du patient utilisées dans ce traitement et les résultats qui en sont issus leur sont accessibles.III.-Les concepteurs d'un traitement algorithmique mentionné au I s'assurent de l'explicabilité de son fonctionnement pour les utilisateurs*».

Adottando l'impostazione basata sui diritti che ha caratterizzato l'intero lavoro, parrebbe potersi affermare che l'utilizzo di sistemi *black-box*, anche qualora le tecniche di *explainable artificial intelligence* forniscano spiegazioni non percepite come sufficienti, possa risultare accettabile nei contesti in cui le prestazioni dell'intelligenza artificiale garantiscano un livello di protezione della salute individuale non raggiungibile altrimenti, e non impongano, al contempo, il sacrificio della salute di altri. È il caso principalmente dei sistemi utilizzati a supporto della diagnosi, spesso basati sul riconoscimento di immagini, che permettono l'identificazione precoce di gravi patologie[981]. In tali situazioni, il tenore dei valori coinvolti, e la possibilità di una tutela più efficace della vita stessa di molti pazienti, porta a sacrificarequasi totalmente il diritto alla spiegazione, in esito al bilanciamento. Rimane presente il rischio di errori nei risultati di tali sistemi e di *bias*con esiti spiacevoli. Un caso tristemente banale, ma già verificatosi, riguarda la differenza, in termini di qualità dei risultati, di un algoritmo impiegato per il riconoscimento del melanoma addestrato con dati relativi a soggetti con la pelle bianca e poi utilizzato su pazienti di colore[982]. Queste potenziali conseguenze negative dovranno essere mitigate dal medico umano – ecco il significato ultimo del

[981] Cfr. ancora, *ex multis*, A. HEKLER ET AL., *Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images*, in *European Journal of Cancer*, 118, 2019, p. 91 ss.; P. RAJPURKAR ET AL., *CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning*, 2017, arXiv:1711.0522; A. HEKLER ET AL., *Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images*, in *European Journal of Cancer*, 118, 2019, p. 91 ss.; A. ESTEVA, B. KUPREL, R. A. NOVOA, J. KO, S. M. SWETTER, H. M. BLAU, *Dermatologist-level classification of skin cancer with deep neural networks cit.*

[982] Cfr. ad esempio L.N. GUO, M.S. LEE, B. KASSAMALI, C. MITA, V.E. NAMBUDIRI, *Bias in, bias out: underreporting and underrepresentation of diverse skin types in machine learning research for skin cancer detection—A scoping review*, in *Journal of the American Academy of Dermatology*, 87, 1, 2022; P. AGGARWAL, F.A. PAPAY, *Artificial intelligence image recognition of melanoma and basal cell carcinoma in racially diverse populations*, in *Journal of Dermatological Treatment*, 33, 4, 2022, p. 2257-2262, doi:10.1080/09546634.2021.1944970.

requisito dello *human in the loop* – ma l'esame globale degli interessi in gioco fa considerare imprescindibili tali tecnologie.

Cercando di ragionare in termini più generali, in questi casi la *tecnicità* della scelta sembra prevalere sulla sua *eticità*, poiché la gerarchia di valori in cui il sistema svolge la valutazione, con precisione pari o superiore all'essere umano, risulta ben chiara: a prevalere su altre istanze, come la piena comprensibilità della tecnologia, è la tutela della vita e dell'integrità fisica dei pazienti. I possibili effetti avversi sono da considerarsi manifestazioni patologiche del funzionamento del sistema, cui cercare di porre rimedio con un controllo umano effettivo, da parte di un sanitario adeguatamente formato sul punto[983]. L'indicazione elaborata dallo strumento si inserisce in un quadro etico chiaro, e replica la valutazione tecnico-specialistica, basata su esperienza e conoscenza, di un sanitario umano.Da questo punto di vista, uno degli interrogativi principali sarà garantire che il quadro regolatorio in materia di intelligenza artificiale che si sta sviluppando non limiti in modo irragionevole tali sistemi, imponendo requisiti tecnologicamente insostenibili, come potrebbero rivelarsi alcune delle previsioni, già analizzate, dell'art. 14 della Proposta di Regolamento europeo[984]. Anche la recente novella del *Code de la santé publique* francese pare poter portare all'esclusione di applicazioni dell'intelligenza artificiale la cui utilità sociale sembra indiscussa, se verrà interpretata rigidamente dal personale sanitario e dalle autorità di regolazione.

Il quadro muta, invece, quando l'intelligenza artificiale sia impiegata in scelte che attengono, in tutto in parte, alla gestione di risorse sanitarie scarse, in cui a prevalere è l'*eticità* della scelta. In tali situazioni, infatti, l'uso della tecnologia non avviene a valle di un chiaro bilanciamento di valori, ma per svolgere scelte – come l'ammissione o meno a un reparto di terapia intensiva – che, pur basate sull'analisi di parametri tecnici, hanno come oggetto essenziale proprio la gerarchizzazione di diritti, interessi e principi etici. Definire le priorità nell'accesso a una terapia intensiva significa prima di tutto determinare che il diritto alla vita e all'integrità fisica di qualcuno prevarrà su quello di qualcun altro.Il prevalente contenuto etico di queste scelte non è privo di conseguenze. In primo luogo, come già accennato, la decisione tragica non potrà mai apparire completamente giusta al decidente e alla società, per quanto raffinate siano le valutazioni predittive su cui si basi. Essa, infatti, consiste sempre nel sacrificio di un interesse in favore di un altro, in un contesto in cui l'unica opzione che appare veramente moralmente accettabile è la protezione di entrambi. Ipotesi di delega pressochè totale alla tecnologia di questo tipo di scelte devono, allora, essere esaminate con

[983] Sottolineano la necessità di un'adeguata formazione del personale medico incaricato di interfacciarsi con sistemi intelligenti, tra gli altri, E.R. HAN ET AL., *Medical education trends for future physicians in the era of advanced technology and artificial intelligence: an integrative review*, in *BMC Medical Education*, 19, 1, 2019; S.A. WARTMAN, C.D. COMBS, *Medical education must move from the information age to the age of artificial intelligence*, in *Academic Medicine*, vol. 93, 8, 2018.

[984] Si rimanda all'analisi svolta *supra*, p. 224 ss.

particolare cautela. Esse, infatti, non sono riducibili a semplici valutazioni di efficienza, e riguardano i momenti fondamentali della malattia e della morte. In analogia con quanto affermato in materia di decisione giudiziaria, affidare a sistemi intelligenti scelte di questo tipo appare disumanizzante, poiché tali strumenti, non partecipando della condizione umana, non hanno esperienza di tali momenti[985]. In secondo luogo, la possibilità di *bias* ed esiti discriminatori non sembra sempre bilanciabile con l'efficienza. I sistemi di supporto alla diagnosi clinica già visti forniscono un'indicazione che – andando oltre le capacità dell'essere umano – può portare a una diagnosi precoce a garanzia della salute di un singolo paziente. Come già detto, pericoli connessi a eventuali disfunzioni sono sicuramente rilevanti, ma il modo più corretto di affrontarli pare da cercare nella sorveglianza di un medico umano dotato della formazione adeguata. Il controllo e la formazione del personale sanitario potrebbero rivelarsi fondamentali anche nel caso di tecnologie di intelligenza artificiale applicata a scelte tragiche, ma la diversità di tali sistemi impone maggiore cautela: in tali casi, infatti, l'indicazione non è finalizzata al potenziale benessere di un singolo paziente, ma, per definizione, al benessere di uno a discapito di altri. Inoltre, come già osservato, non esiste opzione genuinamente giusta, esprimibile in termini di efficienza. La stessa identificazione di eventuali esiti discriminatori è molto più complessa, perché il confine con criteri di priorità che potrebbero considerarsi legittimi può risultare piuttosto labile e meritevole di una discussione, in senso lato politica, che non pare facilmente delegabile alla macchina. La citata vicenda delle Linee Guida SIAARTI al principio della pandemia di Covid-19 offre un buon esempio di questa circostanza.

Per queste ragioni, nei contesti tragici, caratterizzati dalla necessità di allocare risorse sanitarie scarse, anche salvavita, sembra necessario predisporre robuste garanzie affinchè la decisione rimanga effettivamente in mani umane. In particolare, non sembra accettabile l'impiego di tecnologie non accompagnate da solide strategie per l'elaborazione di una spiegazione: l'indicazione da queste ultime fornita al medico umano dev'essere priva di zone d'ombra, in modo che questo possa valutarla serenamente come un elemento in più su cui basare la decisione finale, insieme alla sua esperienza e alle sue competenze tecniche. In mancanza di strategie di *explainable artificial intelligence*applicabili ed efficaci, dovrebbe rinunciarsi all'utilizzo di sistemi non conoscibili, anche qualora promettano livelli maggiori di efficienza. Ciò anche alla luce della difficoltà di definire che tipo di soluzione del dilemma tragico sia preferibile e più efficiente senza un'approfondita valutazione di carattere, come già detto, intrinsecamente etico. In breve, adottando la tassonomia di nuovi diritti teorizzata in questo lavoro, non sembrano esserci in gioco valori che impongono di ridurre l'ambito applicativo dei diritti alla spiegazione e al controllo umano sul

[985] Sul punto cfr. l'ampia analisi svolta *supra*, p. 256 ss., e la letteratura ivi citata.

sistema.Anzi, la conservazione della decisione nelle mani di un'effettivo *human in the loop* – in grado di decidere se usare o meno il sistema, comprenderne il funzionamento e il risultato, analizzare le variabili determinanti per la sua elaborazione -sembra necessaria a garantire al meglio diritti e interessi coinvolti nella decisione tragica, in primo luogo dignità, vita e integrità fisica.Da questo punto di vista, la già analizzata formulazione rigorosa adottata dall'art. 14 della Proposta di Regolamento dell'Unione Europea in materia di intelligenza artificialepotrebbe rivelarsi un valido strumento di regolazione, qualora venisse approvata nella formulazione corrente. I requisiti ivi previsti, infatti, pur lasciando spazio a interpretazioni più liberali, sembrano orientare lo sviluppo di sistemi intelligenti verso la direzione appena descritta. Anche la recente novella del *Code de la santé publique* francese pare essere stata concepita principalmente al fine di porre limiti all'utilizzo di sistemi autonomi in questo genere di decisioni eticamente sensibili. Si sono già passate in rassegna, in diversi punti del lavoro, le critiche che è possibile rivolgere a queste norme, considerate eccessivamente rigide e accusate da più parti di essere tecnologicamente insostenibili. È significativo, ad ogni modo, notare come esse risultino particolarmente calzanti in un ambito, le scelte tragiche, che rappresenta di certo uno scenario estremo. Ragionando a contrario, può ricavarsene un'indicazione ulteriore di come tali disposizioni, in particolare l'art. 14 della Proposta di Regolamento, per la sua portata generale, rischino di non risultare adatte a ogni contesto e rallentare, in modo non sempre necessario, lo sviluppo tecnologico.

Tanto detto, rimane da analizzare se i principi visti finora, parsi adatti a una possibile regolazione generale dell'utilizzo di tecnologie intelligenti nell'allocazione di risorse sanitarie limitate, risultino applicabili anche quando la situazione di estrema scarsità sia repentina e non prevista, perché causata dal presentarsi di un'emergenza. Quest'ultima, infatti, riduce il tempo, le energie e le risorse umane a disposizione, obbligando i sanitari a un numero elevato di scelte tragiche, da compiersi rapidamente, in condizioni di stress estremo e, talvolta, sulla base di informazioni riguardanti i pazienti solo parziali. Nonostante gli ordinamenti democratici, tipicamente, rifiutino di considerare l'emergenza fonte autonoma di diritto, essa, di certo, restringe l'insieme delle soluzioni praticabili, obbligando a modalità operative che non sarebbero accettabili in tempi ordinari[986]. Può accadere,

[986] Dei confini della legislazione d'emergenza ha avuto modo di occuparsi anche la Corte costituzionale italiana, nella notissima sentenza 15 del 1982, riguardante l'estensione dei termini di custodia cautelare per i delitti attinenti all'emergenza connessa all'estremismo politico che l'Italia viveva in quel periodo, disposta con il D.L. n. 625 del 1979. La pronuncia, pur riconoscendo il dovere del Legislatore di fronteggiare con misure straordinarie eventuali situazioni emergenziali, ha chiarito che tali misure non si collocano mai totalmente *extra ordinem*, e sono anzi vincolate a limiti precisi, pena la loro incostituzionalità: esse, infatti, devono essere limitate nel tempo (l'emergenza permanente non è più tale), ispirarsi a criteri di necessità e proporzionalità, rispettare, in ogni caso, il nucleo essenziale, non comprimibile, dei diritti fondamentali. Considerazioni in parte accostabili, pur nella radicale diversità delle situazioni, sono state formulate anche dalla Corte Suprema degli Stati Uniti, nelle sentenze del 2004 *Rasul v. Bush* e *Hamdi v. Rumsfeld*, che hanno riconosciuto l'incomprimibilità del diritto all'*habeas corpus* dei cittadini stranieri detenuti nel carcere di Guantanamo. In diversi casi, inoltre, sono le stesse costituzioni democratiche a disciplinare forme e limiti dei poteri d'emergenza: è il caso, ad esempio, della Costituzione spagnola (art. 116) e della Costituzione francese (art. 16). In letteratura si rimanda,

allora, che decisioni tragiche siano assunte sulla base di rudimentali stime probabilistiche di efficacia di un trattamento, o valutazioni interamente soggettive d'altro genere, basate su un limitato numero di esperienze del singolo curante. La ricerca di coerenza interna, pur all'interno della situazione emergenziale, ed esigenze di rapidità possono portare a scelte basate in misura eccessiva su automatismi potenzialmente fallibili e dall'aspetto talvolta arbitrario, come la fissazione di un limite d'età per l'accesso a determinati trattamenti, o l'applicazione meccanica, trascurando ogni altro elemento, di quozienti numerici come il SOFA score.Si tratta di quanto avvenuto in vari paesi, Italia compresa, in occasione del diffondersi del virus Sars-Cov-2[987].

In scenari di questo genere, l'impiego di strumenti di intelligenza artificiale per la gestione di tali decisioni pare tendenzialmente da accogliere con favore.Gli algoritmi, infatti, sono insensibili a fatica e stress, e operano rapidamente: la qualità delle loro prestazioni, al contrario di quanto accade per i sanitari umani, non è condizionata dalla situazione di emergenza. Le loro indicazioni, sempre basate sull'analisi di grandi moli di dati, potrebbero risultare particolarmente utili in contesti in cui il tempo che il medico può dedicare alla singola decisione, e quindi all'esame delle informazioni a sua disposizione, è particolarmente scarso. D'altro canto, il rischio di distorsione dell'automazione, in simili casi, sembra particolarmente alto, e le possibilità di identificare eventuali malfunzionamenti nei risultati dei sistemi molto basse. In poche parole, il pericolo di una delega nei fatti totale alla tecnologia della scelta tragica appare più che concreto. Queste circostanze dovrebbero portare alla massima cautela verso l'utilizzo di tecnologie non completamente spiegabili anche in tali situazioni, ribadendo le argomentazioni viste poco sopra. Tuttavia, le specificità dell'emergenza – e la forzata modifica a ciò che risulta moralmente accettabile, e legalmente scusato, che implica – non possono non considerarsi, e sistemi intelligenti basati sull'apprendimento automatico potrebbero contribuire a rendere più razionali scelte il cui processo decisionale è particolarmente deteriorato. Probabilmente, allora, i protocolli normativi per la

---

tra i molti e senz'animo di completezza, a B. ACKERMAN, *The emergency constitution*, in *The Yale Law Journal*, 113, 5, 2004, p. 1029-1091e*Before the next attack. Preserving civil liberties in an age of terrorism*, New Haven (US), 2007; P. BONETTI, *Terrorismo, emergenza e costituzioni democratiche*, Bologna, 2006; G. DE MINICO, *Costituzione, emergenza e terrorismo*, Napoli, 2016; G. GALLI, *Difesa dell'imputato e speditezza del processo: dalla Costituzione alle leggi dell'emergenza*, Milano, 1982; O. SPATARO, *Stato di emergenza e legalità cosituzionale alla prova della pandemia*, in *Federalismi.it*, 11, 2022, p. 158-186; A. D'ALOIA, *Costituzione ed emergenza. L'esperienza del coronavirus*, in *BioLaw Journal – Rivista di BioDiritto*, Special Issue 1, 2020, p. 7-12; R. RAVÌ PINTO, *Brevi considerazioni su stato di emergenza e diritto costituzionale*, in *BioLaw Journal – Rivista di BioDiritto*, Special Issue 1, 2020, p. 43-50.

[987]Sul tema cfr., ad esempio, B. HERREROS, P. GELLA, D.R. DE ASUA, *Triage during the Covid-19 epidemic in Spain: better and worse ethical arguments*, in *Journal of Medical Ethics*, 46, 7, 2020; R.D. TRUOG, C. MITCHELL, G.Q. DALEY, *The Toughest Triage — Allocating Ventilators in a Pandemic*, in *The New England Journal of Medicine*, 382, 2020, p. 1973-1975; E. KUCEWICZ-CZECH, M. DAMPS, *Triage during the Covid-19 pandemic*, in *Anaesthesiology Intensive Therapy*, 52, 4, 2020, p. 312-315; M. IMARISIO, *Coronavirus, il medico di Bergamo: «Negli ospedali siamo come in guerra. A tutti dico: state a casa»*, Corriere della Sera, 9 marzo 2020; P. ARENSI, *"I giorni più bui del Covid in ospedale? È stato come finire per caso in guerra"*, Il Giorno, 31 maggio 2022; *The tough ethical decisions doctors face with covid-19*, The Economist, 2 aprile 2020.

pianificazione della gestione di situazioni emergenziali – la cui inadeguatezza in molti paesi, in occasione dellapandemia di Covid-19, è stata, del resto, commentata da più parti[988] - non dovrebbero precludere l'uso di tali strumenti ausiliari da parte del personale sanitario, a prescindere dalla loro eventuale opacità, anche nei sistemi normativi in cui, per ipotesi, il loro utilizzo in circostanze ordinarie fosse proibito.

Allo stesso tempo, deve sottolinearsi un rischio: l'emergenza, e la necessità di scelte tragiche che porta con sé, dev'essere frutto di circostanze transeunti e non prevedibili, e non di un'errata pianificazione delle risorse. Scelte generali riguardanti la spesa sanitaria, pubblica e privata,possono portare alla scarsità di presidi sanitari anche in situazioni ordinarie, o di circostanze che portano al sovraccarico dei sistemi ospedalieri del tutto prevedibili[989]. L'importanza delle decisioni di pianificazione è stata evidenziata anche dal modo diseguale con cui i sistemi sanitari dei paesi del mondo sono stati in grado di fronteggiare l'epidemia causata dal virus Sars-Cov-2. In poche parole,le *first-order choice* – riprendendo la terminologia di Calabresi e Bobbit – sono scelte etiche, e rimangono il primo fattore determinante la necessità di operare, in un momento futuro, le *second-order choice*. L'utilizzo di tecnologie a supporto di queste ultime, e il guadagno in termini di efficienza generale – e, specialmente in situazioni emergenziali,di quantità delle vite salvate - che ne può derivare non può servire a nasconderlo. Né la maggior efficacia garantita dalla tecnologia può fungere da alibi per un'eventuale riduzione degli investimenti sanitari complessivi, una *first-order choice* che rimarrebbe politica ed etica, e non meramente tecnica. La difficoltà a percepire la tragicità delle scelte d'indirizzo generale – quando non la loro stessa esistenza – era, del resto, stata rilevata già dagli Autori di *Tragic choices*, come già riportato nel corso di questo lavoro[990]. Un impiego dell'intelligenza artificiale rispettoso della centralità dell'essere umano deve evitare che la tecnologia, grazie all'apparenza di oggettività che spesso la circonda, acuisca tale difficoltà, specialmente in un ambito intrinsecamente delicato come quello medico.

[988]Cfr. ad esempio C. ALFIERI, M. EGROT, A. DESCLAUX, K. SAMS, *Recognising Italy's mistakes in the public health response to COVID-19*, in *Lancet*, 399, 10322, 2022, p. 357-358; UK GOVERNMENT, *Uk pandemic preparedness*, Policy Paper, 5 novembre 2020, https://bit.ly/3gbIIRa (15 ottobre 2022);M. GASPERETTI, *Coronavirus, il piano anti-pandemia che l'Italia non ha seguito*, Corriere della Sera, 28 marzo 2020; D. DE FELICE, *Il piano pandemico inadeguato e obsoleto. Ne ho parlato Francesco Zambon*, il Fatto Quotidiano, 7 luglio 2021; G. TREMLET, *How did Spain get its Coronavirus response so wrong?*, The Guardian, 26 marzo 2020.

[989]Notizie di situazioni critiche nella gestione di alcuni presidi terapeutici e ospedali al limite della saturazione, del resto, non mancavano nelle cronache precedenti alla pandemia di Covid-19, v. ad esempio S. RAVIZZA,*Milano, terapie intensive al collasso per l'influenza: già 48 malati gravi, operazioni rinviate*, Corriere della Sera, 10 gennaio 2018; *Influenza, ospedali in tilt fra tagli e psicosi vaccino: "Lorenzin arrivata tardi"*, il Fatto Quotidiano, 24 gennaio 2015; A. MACMILLAN, *Hospitals Overwhelmed by Flu Patients Are Treating Them in Tents*, Time, 18 gennaio 2018; S. KARAMANGLA, *California hospitals face a 'war zonÈ of flu patients — and are setting up tents to treat them*, Los Angeles Times, 16 gennaio 2018; R. SALAMANCA, *La gripe colapsa los hospitales de media España*, El Mundo, 12 gennaio 2017; B. COLLET, *Epidémie de grippe: «En dix-neuf ans de métier, je n'ai jamais vu ça»*, le Monde, 12 gennaio 2017.

[990]Cfr. G. CALABRESI, P. BOBBIT, *Tragic choices cit.*, p. 20 ss.

Conclusioni

Questo lavoro, come si è visto, è organizzato in tre parti. La prima indaga le principali caratteristiche tecniche delle tecnologie intelligenti, ripercorre le tappe fondamentali del loro sviluppo dal punto di vista storico, ricostruisce nelle linee essenziali il dibattito filosofico e sociologico sulla rivoluzione tecnologica e le principali prospettive di una sua regolazione. La seconda e la terza, più corpose, affrontano il tema centrale dello studio: l'impatto dell'intelligenza artificiale sui diritti fondamentali. La seconda parte, in particolare, si concentra sulle sfide poste da questa famiglia di tecnologie variegata e complessa ad alcune delle garanzie giuridiche fondamentali consolidate nella tradizione democratica, e sulle possibili strategie per farvi fronte. Si tratta, in particolare, dei diritti afferenti alla sfera dell'identità personale, della libertà di manifestazione del pensiero, del principio di eguaglianza. La terza parte, invece, è dedicata alla possibilità di aggiornare il catalogo dei diritti fondamentali con nuove posizioni giuridiche, la cui tutela è resa necessaria dai cambiamenti provocati dall'avvento dell'intelligenza artificiale: il diritto a conoscere la natura artificiale di un sistema, il diritto a una spiegazione dei risultati di quest'ultimo, il diritto a controllo e sorveglianza umani sulle tecnologie avanzate. Questi *nuovi diritti* sono, poi, analizzati in tre ipotetici ambiti di applicazione, scelti innanzitutto per il rilievo dei diritti fondamentali tipicamente coinvolti: la Pubblica Amministrazione, il sistema giustizia, l'attività medica.

L'opera, comunque, raccoglie solo una piccola parte delle possibili riflessioni riguardanti le ripercussioni sui diritti fondamentali dell'intelligenza artificiale. Inoltre, ogni ricerca sul tema rischia di vedere i propri risultati superati dal progresso tecnologico, che pone, giorno dopo giorno, il diritto di fronte a interrogativi imprevisti. Proprio la radicale novità delle sfide generate dall'avvento di sistemi intelligenti, del resto, ha portato a concludere l'opera col tema dei nuovi diritti. Dunque, i risultati cui il lavoro approda – dei quali si tenta, ora, una sistematizzazione sintetica – devono considerarsi, più che in ogni altro caso, solo il punto di partenza per sviluppi, correzioni e miglioramenti futuri.

I diritti fondamentali e lo "schermo" della tecnica

Preme ribadire, innanzitutto, l'obiettivo fondamentale della ricerca: evidenziare come la rivoluzione connessa all'intelligenza artificiale chiami in causa la categoria giuridica dei diritti fondamentali. La complessità delle tecnologie in questione, infatti, genera il rischio di considerarle interesse esclusivo di un gruppo ristretto di esperti, in possesso delle conoscenze specialistiche per il loro sviluppo, senza che il resto della società sia tenuto ad occuparsene. Si tratta di una percezione del tutto

erronea: in primo luogo, perché le tecnologie intelligenti hanno già invaso molti ambiti della vita quotidiana, e continueranno a farlo in futuro, venendo utilizzate da quasi ogni individuo, a prescindere dalle competenze tecniche; in secondo luogo, perché diverse caratteristiche di tali tecnologieincidono direttamente sulla relazione tra uomo e macchine e tra uomo e natura. Come ampiamente chiarito nel corso dell'opera, è il caso, principalmente, della capacità di portare a termine compiti complessi – come attività valutative e decisionali – in parziale o totale autonomia, e dei limiti che alcune di esse presentano all'interpretabilità del loro funzionamento e alla spiegabilità dei loro risultati. Queste componenti tecniche portano a ridiscutere il rapporto dell'essere umano con la realtà che lo circonda da due punti di vista: mettendo in dubbio la sua natura di agente esclusivo di determinate attività in cui è coinvolta l'intelligenza, in precedenza non automatizzabili, ed erodendo le possibilità di controllo completo sugli artefatti da egli stesso creati, totale prima dell'avvento dell'intelligenza artificiale. Proprio per questo, acutamente, molte delle prime riflessioni etico-giuridiche su questi sistemi hanno posto l'accento sulla necessità di indirizzare lo sviluppo tecnologico verso la direzione di un'IA affidabile – e quindi sottoponibile alla valutazione di un essere umano – e *human-centric*.

Il rischio che la complessità tecnica finisca per nascondere la portata delle questioni di fondo coinvolte è presente anche nella regolazione del fenomeno. Le prime norme in materia, già in vigore o in fase di elaborazione, infatti, non possono che presentarsi come norme tecniche, volte a orientare l'azione degli specialisti che progettano e sviluppano le applicazioni tecnologiche in esame. È il caso di gran parte dei testi normativi commentati nel corso del lavoro. La caratteristica di regole di questo tipo – il cui sviluppo e continuo aggiornamento, nella società industriale, non è certo una novità – è di fornire codici di comportamento anche molto dettagliati, la cui lettura è spesso più agevole agli specialisti del settore di riferimento che ai tecnici del diritto. Enunciazioni di principio e richiami ad altri documenti sono spesso del tutto assenti, o comunque meno frequenti che in testi normativi d'altro tipo. Ciò può portare a oscurare la circostanza che tali norme – nel settore dell'intelligenza artificiale, ma anche in altri campi di applicazione – sono, talvolta, uno strumento fondamentale per l'attuazione dei diritti fondamentali e dei principi generali dell'ordinamento. Le ragioni sono intuibili: si pensi, ad esempio, agli interessi giuridici alla tutela dell'ambiente e della salute collettiva, che trovano la loro effettività, prima di tutto, in regole tecniche in grado di imbrigliare gli operatori industriali più inquinanti, al di là della loro enunciazione per principi generali nelle carte fondamentali, o dell'elaborazione dottrinale in materia. Nel caso dell'intelligenza artificiale, alcuni dei testi normativi in materia dimostrano una certa consapevolezza di quanto appena affermato: è il caso, in particolare, della Proposta di Regolamento in fase di discussione presso le istituzioni europee, che inquadra tra i suoi obiettivi

fondamentali proprio la tutela dei diritti fondamentali. Questo lavoro fa un passo in più e sostiene che quando i valori in gioco riguardano la protezione della sfera della personalità individuale nei confronti di un potere, inteso in senso lato (pubblico o privato, umano o tecnologico), proprio l'elaborazione di norme tecniche può rappresentare un indice della necessità di aggiornare, a livello teorico, il catalogo dei diritti fondamentali con nuove tutele giuridiche di rango primario, a difesa di corrispondenti, nuovi pericoli. Solo raramente, infatti, un nuovo diritto prende vita con l'enunciazione in costituzioni e carte internazionali – e, quando è così, rischia di ridursi a formula vuota, di cui l'ordinamento non garantisce l'effettività – più spesso fa la sua comparsa in disposizioni di dettaglio volte alla sua tutela in uno specifico ambito, nelle sentenze dei giudici e, più tardi, nelle riflessioni della dottrina, avviando un percorso del quale la positivizzazione solenne (peraltro non sempre presente) è solo il termine finale. Una delletesi principali sostenute in quest'opera è che un fenomeno di questo genere stia avvenendo riguardo alla regolazione dell'intelligenza artificiale, in riferimento ai menzionati nuovi diritti a conoscere la natura umana o artificiale di un sistema, alla spiegazione dei risultati della tecnologia, e al controllo umano su quest'ultima.

Molte tecnologie, un solo diritto

Come evidenziato nel corso del lavoro, allo stato dell'arte la regolazione dell'intelligenza artificiale è costituita da un nutrito elenco di documenti di *soft-law*, spesso provenienti da istituzioni internazionali, poche norme di *hard-law* e alcune ipotesi di regolazione allo studio, la principale delle quali è la menzionata Proposta di Regolamento dell'UE sull'intelligenza artificiale. La maggioranza di questi documenti è di taglio estremamente generale, risolvendosi spesso, nel caso degli strumenti di *soft-law*,in dichiarazioni di principi e linee guida. Anche le ipotesi di regolazione attualmente in vigore e la maggioranza di quelle in discussione adottano un approccio ampio, puntando a disciplinare l'intero insieme delle tecnologie riconducibili alla famiglia dell'intelligenza artificiale (è il caso, ad esempio, della Proposta europea) o, comunque, un loro settore esteso (si pensi alla più volte citata *Directive on automated decision-making* canadese). La circostanza è la conseguenza dell'assenza pressoché totale, in molti ordinamenti, di norme applicabili alle peculiarità tecniche dei sistemi intelligenti: la necessità di dettarne una prima regolazione spinge a formulare sistemi di regole onnicomprensivi. Ciò ha sicuramente il pregio di fissare dei punti fermi – anche con norme tecniche rivolte ai pratici della disciplina, come già esposto – validi per ogni possibile applicazione dell'intelligenza artificiale: una caratteristica di sicura importanza dal punto di vista dei diritti fondamentali, la cui tutela, al netto dei bilanciamenti con altri diritti e interessi di volta in volta coinvolti, prescinde ovviamente dal campo di applicazione. D'altro canto, vi sono

ambiti applicativi dell'intelligenza artificiale che presentano peculiarità che non possono essere ignorate. L'ultima parte del lavoro, come già detto dedicata ai principali utilizzi dell'intelligenza artificiale in tre domini specifici – medicina, pubblica amministrazione e giustizia – ha avuto il fine, per l'appunto, di sottolineare tali peculiarità, e le loro ripercussioni in materia di diritti fondamentali. Talune applicazioni delle tecnologie intelligenti in contesti particolarmente delicati per il rango degli interessi coinvolti, forse, necessiterebbero di una disciplina specifica, ritagliata sull'ambito di riferimento. Alcuni ordinamenti meno attenti di quello europeo alla regolazione delle tecnologie avanzate sembrano intenzionati a procedere in questa direzione: è stato citato il caso delle autorità di Regno Unito, Canada e Stati Uniti, che hanno elaborato un documento congiunto (meramente di *soft-law*) contenente alcuni principi generali relativi all'applicazione dell'apprendimento automatico in ambito medico. Non può escludersi, in ogni caso, che anche l'Unione Europea emani, in futuro, discipline di dettaglio rivolte ad ambiti specifici, magari poggiando sul quadro generale introdotto, in caso di approvazione, dalla Proposta di Regolamento attualmente allo studio delle istituzioni.

Nuove sfide, vecchi diritti e istituti giuridici da ripensare

La seconda parte del lavoro, come già riportato anche in queste conclusioni, è stata dedicata alle conseguenze dell'avvento dell'intelligenza artificiale su tre istanze giuridiche fondamentali: il diritto all'identità personale (nelle sue composite declinazioni); il diritto alla libera manifestazione del pensiero; l'effettività del principio di eguaglianza. L'analisi delle situazioni giuridiche afferenti all'identità personale e delle ripercussioni dei sistemi intelligenti sulla protezione della libertà d'espressione ha fatto emergere in modo particolare i limiti di due categorie giuridiche considerate, fino ad oggi, fondamentali per la regolazione delle nuove tecnologie: la concezione della protezione dei dati personali come diritto individuale, e del consenso come principale condizione di liceità del trattamento; il principio della *liability exemption* per gli intermediari di internet.

L'inquadramento della protezione dei dati come diritto fondamentale caratterizza, con sfumature differenti, gli ordinamenti democratici occidentali, ed è patrimonio, senza ambiguità, in primo luogo dei sistemi europei e delle norme eurounitarie sul tema. Non si tratta di certo di un modello da abbandonare; allo stesso tempo, però, non può non riconoscersi che troppo spesso esso si risolve in una tutela solo apparente. La pervasività della sfera digitale e la frequenza delle interazioni quotidiane di ogni individuo con operatori di servizi internet, infatti, rendono inefficace un modello che identifica nel consenso (all'utilizzo di propri dati, o al negozio giuridico per cui esso è necessario) la condizione di legittimità privilegiata del trattamento di dati personali. È a tutti noto il fenomeno dell'approvazione meccanica, di fatto inconsapevole, di trattamenti accompagnati da

informative lunghe e complesse, il cui contenuto – al di là degli ipotetici tempi di lettura – sarebbe comunque difficilmente accessibile per la grande maggioranza degli utenti, del tutto digiuna della materia. A ciò si aggiungono i trattamenti di dati che l'utente accetta con l'utilizzo di *social network* e piattaforme, resi leciti dalle condizioni contrattuali di queste ultime, unilateralmente predisposte senza margine di negoziazione e, nella larga maggioranza dei casi, approvate senza prenderne visione. Inoltre, come analizzato nel dettaglio nel corso dell'opera, gli sviluppi connessi all'avvento dell'intelligenza artificiale rendono particolarmente complesso prevedere le reali conseguenze di un trattamento dati, innanzitutto riguardo alla possibilità di ricavare, a partire da dati determinati, informazioni ulteriori sulla persona. Le possibili strategie per restituire effettività al diritto al controllo sui dati personali passano per la valorizzazione della sua dimensione superindividuale. In primo luogo, con norme di diritto pubblico che impediscano, almeno in parte, trattamenti di dati non necessari alle piattaforme, intervenendo, a tutela della collettività dei loro utenti, sull'assetto contrattuale da esse disegnato, in analogia a quanto accade in altri settori caratterizzati da una marcata asimmetria tra le parti, come il diritto del consumatore. In secondo luogo, valorizzando forme di gestione superindividuale in materia di dati personali, in grado di favorire un approccio più informato, consapevole e razionale al tema dei dati, anche attraverso l'impiego di specialisti del settore. Tali enti collettivi – spesso significativamente chiamati *data trust* – potrebbero rappresentare gruppi di individui anche molto numerosi nella negoziazione coi grandi operatori di internet, in analogia a quanto avviene in materia di contrattazione collettiva nel diritto del lavoro, rompendo lo schema, attualmente dominante, che vede questi ultimi predisporre, in totale autonomia, le condizioni dei loro servizi.

Volgendo lo sguardo al principio della *liability exemption* per gli intermediari di internet, deve rilevarsi, in primo luogo, che esso è – tenendo conto della novità dell'oggetto di regolazione – relativamente risalente, risultando presente in vari ordinamenti, innanzitutto quello dell'UE e quello statunitense, fin dagli anni '90. Il principio ha rappresentato una delle basi fondamentali dello sviluppo dell'internet interattivo, liberando le piattaforme dell'onere di un controllo generalizzato e capillare sui contenuti in esse diffusi dagli utenti, del resto nella pratica irrealizzabile. L'evoluzione subita dal mercato dei servizi internet è nota, ed è stata analizzata nel dettaglio nel corso del lavoro: pochi grandi operatori hanno acquisito un ruolo dominante sul mercato, accumulando poteri di controllo del discorso pubblico inediti in mani private. Il web, così, ha smesso da tempo di somigliare all'*agorà* digitale in cui tutti avrebbero avuto accesso a inedite possibilità di esprimersi che prometteva di essere ai suoi inizi, per lasciare spazio al proliferare a volte incontrollato di notizie false e discorsi d'odio, da un lato, e a controversi e sempre più frequenti episodi di censura privata da parte delle piattaforme, dall'altro. Per queste ragioni, il principio della *liability exemption*

è stato, in tempi recenti, messo fortemente in discussione, e accusato di essere uno strumento giuridico ormai inadatto alla regolazione di un fenomeno complesso come l'esercizio della libertà d'espressione su internet. Nonostante l'ampio dibattito in materia, l'esenzione, in realtà, non pare totalmente da abbandonare: gravare le piattaforme dell'obbligo di svolgere un controllo generalizzato di liceità sui contenuti generati e condivisi dagli utenti le porterebbe, prevedibilmente, ad assumere un atteggiamento censorio estremamente rigoroso, al fine di non incorrere in possibili sanzioni, con ogni conseguenza anche riguardo al c.d. *chilling effect* tra gli utenti. La soluzione, quindi, non è da ricercarsi nella revisione totale del principio, ma in un suo aggiornamento che tenga conto della crescente complessità delle attività che gli operatori svolgono sui contenuti, prima di tutto in materia di indicizzazione e personalizzazione in funzione delle informazioni raccolte sull'utente. Risulta sempre più complesso, in molti casi, affermare che motori di ricerca e reti sociali siano meri intermediari e – ferma l'impraticabilità e non auspicabilità di un dovere di controllo generalizzato – pare opportuno adeguare l'ordinamento all'evoluzione delle loro attività e alla posizione di "luoghi privilegiati" per l'esercizio della libertà d'espressione che hanno assunto. Essi, allora, dovrebbero essere spinti, con appositi strumenti normativi di *hard law* – trovare le strategie più adatte, come a breve si dirà, è una delle sfide principali del costituzionalismo contemporaneo – a mettere in atto ogni misura tecnica in grado di garantire che, al loro interno, da un lato non proliferino in modo incontrollato contenuti illeciti, dall'altro non avvengano forme ingiustificate di censura privata, magari automatizzate. L'impiego di tecnologie intelligenti, ovviamente, dovrebbe essere coinvolto in una regolazione di questo tipo, posto il suo ruolo insostituibile nell'attività di moderazione. Si tratta, peraltro, di obiettivi che sembrano in parte ispirare alcune recenti ipotesi di regolazione, in primo luogo il *Digital Service Act* approvato di recente dal Parlamento Europeo, estensivamente commentato nel lavoro.

L'intelligenza artificiale nei processi decisionali e le sue altre applicazioni

Il lavoro ha dedicato un'attenzione particolare all'utilizzo dell'intelligenza artificiale in procedimenti decisionali e valutativi, a supporto o in sostituzione integrale del decisore umano. In particolare, nella seconda parte sono state analizzate le strategie di *nudging* algoritmico impiegate nei servizi di internet interattivo al fine di influenzare scelte, gusti e opinioni degli utenti, e le loro possibili ripercussioni per la libera formazione dell'identità personale, la libertà d'espressione, la stessa tenuta del sistema democratico. Anche l'analisi dei possibili nuovi diritti dell'era dell'intelligenza artificiale condotta nella terza parte ha trattato in modo esteso il contesto della decisione, ritenendo che le menzionate esigenze di conoscibilità dei sistemi, controllo su di essi e spiegazione dei loro output vi trovassero un campo di applicazione privilegiato. È innanzitutto di

fronte a una decisione che lo riguardi, infatti, che l'individuo ha interesse a conoscere natura e identità del decisore, ragioni della scelta e, in caso di una sua automazione, l'intensità della sorveglianza su di essa dell'essere umano, al fine di poterne esaminare e contestare l'esito, e sapere a chi rivolgersi per farlo.

L'impatto dell'intelligenza artificiale sui meccanismi decisionali – dei quali l'uomo perde, in molti casi, il monopolio – è testimoniato anche dalla particolare attenzione riservata dal diritto alla loro automazione. Il tema è trattato copiosamente dalla letteratura giuridica, ed è stato già oggetto di regolazione in diversi ordinamenti: si pensi all'art. 22 del GDPR, alla *Directive on automated decision-making* canadese, o alle norme in materia di decisione automatizzata introdotte nell'ordinamento francese con recenti modifiche alla *Loi informatique et libertés* e alla *Loi de bioéthique*. I limiti di tali normative, allo stato dell'arte, sono già stati evidenziati nel corso del lavoro: l'ambito applicativo estremamente circoscritto – la normativa europea disciplina le sole decisioni interamente automatizzate, quelle canadesi e francesi si rivolgono ad ambiti di applicazione definiti – e l'estrema genericità della loro formulazione. Inoltre, per quanto l'importanza della normativa in materia di protezione e trattamento dei dati personali per la disciplina dell'intelligenza artificiale sia già stata evidenziata, anche in queste conclusioni, preme ribadire che essa, nella regolazione dei sistemi decisionali automatizzati, si espone al rischio di rilevanti vuoti di tutela. Infatti, l'elaborazione di grandi moli di dati anonimizzati (e dunque, nel caso europeo, fuori dal campo di applicazione della disciplina sui dati personali) permette comunque lo sviluppo di modelli che, una volta calati sulle caratteristiche di un determinato individuo, rendono possibile formulare profilazioni e predizioni del comportamento estremamente accurate.

Più in generale, l'attenzione giustamente riservata dal diritto al problema della decisione – e in particolar modo della decisione riguardante l'individuo, con la costruzione di relative garanzie in capo a quest'ultimo – rischia di portare a trascurare altri utilizzi dell'intelligenza artificiale applicata all'analisi dei dati estremamente rilevanti, e il cui potenziale impatto sui diritti individuali, pur indiretto, pare ugualmente allarmante. Si tratta di sistemi in senso lato decisionali, il cui impiego su larga scala, e non per decisioni puntuali sul singolo, sembra poter rendere in larga parte inapplicabili le tutele individuali appena citate. Il frequente utilizzo di dati anonimizzati nel loro sviluppo, o di dati costituiti da informazioni non riguardanti persone fisiche, completa il quadro, escludendo l'applicazione delle principali normative in materia di dati personali, in primo luogo quella europea. Il riferimento è ad algoritmi utilizzati a supporto di decisioni di programmazione e pianificazione del comportamento di organizzazioni complesse, compresi i poteri pubblici. Un'impresa multinazionale, così, potrebbe affidarsi a previsioni assistite dalla tecnologia per

determinare il numero ideale di lavoratori di cui necessita, con le prevedibili conseguenze, anche ad anni di distanza, sul piano occupazionale; governi ed enti pubblici potrebbero determinare le modalità di spesa più efficienti per la fornitura di determinati servizi, anche essenziali, sulla base di stime elaborate con l'ausilio di algoritmi sulla base dei dati di consumo del passato e delle caratteristiche dell'utenza di riferimento, esponendo i cittadini al rischio di disservizi in caso di errori nei dati o aumenti del fabbisogno di determinate risorse dettati dal verificarsi di circostanze anomale. Non si può tacere, inoltre, l'eventualità che l'utilizzo di algoritmi, magari al fine dichiarato di ottenere maggiore efficienza e qualità nei servizi erogati, celi la volontà di ridurre la spesa connessa a determinate prestazioni. Si tratta, in sostanza, dei rischi connessi all'ingresso dell'intelligenza artificiale nelle scelte denominate *first-order choices* da Guido Calabresi e Philip Bobbitt nel loro fondamentale libro *Tragic Choices*, analizzato nell'ultima parte dell'opera. Le *first-order choices* sono le decisioni volte a definire le politiche di allocazione generale delle risorse che generano le condizioni di scarsità alla base delle *second-order choices*, le decisioni atomistiche su quale individuo avrà accesso a una determinata risorsa scarsa. Come già detto, un approccio alla regolazione dell'intelligenza artificiale basato sulle norme in materia di protezione dei dati personali rischia di rivelarsi inefficace per garantire i diritti individuali di fronte all'impiego di tecnologie avanzate in tali contesti. Si tratta, peraltro, di un problema di cui la Proposta di Regolamento in materia di intelligenza artificiale allo studio delle istituzioni europee sembra in grado di farsi carico, almeno parzialmente. Infatti, alcune delle norme da essa previste, in particolare in materia di qualità dei *dataset* di allenamento, sono volte a favorire lo sviluppo di sistemi predittivi e valutativi in grado di fornire risultati di qualità e in cui il rischio di *bias* sia il più possibile ridotto, a prescindere dal coinvolgimento di dati personali o dal loro utilizzo in attività decisionali.

Infine, l'attenzione della dottrina giuridica verso il tema della decisione algoritmica rischia di oscurare i pericoli connessi a tutte le applicazioni dell'intelligenza artificiale non impiegate per l'automazione di attività valutative o decisionali, nemmeno in senso lato. Vengono in gioco, in primo luogo, le tecnologie avanzate impiegate in contesti caratterizzati da un ineliminabile rischio intrinseco, tollerato dall'ordinamento giuridico in ragione dell'utilità sociale che essi generano. L'ovvio esempio, già citato nel corso del lavoro, è rappresentato dalla grande robotica industriale. Si tratta di ambiti già sottoposti a regolazioni corpose, volte a razionalizzare l'ineliminabile livello di pericolo che li caratterizza – nel caso dei sistemi automatizzati impiegati in ambienti produttivi ibridi, ad esempio, a venire in esame è la normativa in materia di sicurezza sul lavoro. Tali testi normativi, nonostante la loro completezza e il volume delle norme di dettaglio che spesso li correda, nella maggioranza dei casi sono stati concepiti prendendo a riferimento lo stato dell'arte tecnologico precedente all'avvento dell'intelligenza artificiale. Ne deriva che, talvolta, potrebbero

rivelarsi lacunosi di fronte alle questioni specifiche sollevate da quest'ultima: la qualità dei dataset di addestramento, da cui possono derivare, nel contesto della robotica industriale, errori di funzionamento di determinati macchinari estremamente pericolosi per i lavoratori impiegati accanto ad essi, può essere, ancora una volta, un valido esempio. Inoltre, adottando nuovamente l'approccio basato sui nuovi diritti teorizzati in questo lavoro, risulta immediatamente evidente come, negli ambienti produttivi in esame, il tema del livello minimo di sorveglianza e intervento umano nel funzionamento di tali nuove tecnologie acquisisca centrale rilevanza.

Il diritto costituzionale di fronte ai nuovi poteri

Come già evidenziato, lo sviluppo tecnologicoha generato un inedito accentramento di potere in poche mani private. L'applicazione dell'analisi dei dati a servizi di internet interattivo e IoT, in particolare, ha reso conveniente l'accumulo di una mole sempre maggiore di dati personali in capo a un numero ristretto di grandi aziende, in grado di migliorare costantemente i propri servizi attraverso la profilazione, sempre più accurata, dei loro utenti. Le sfide poste da questa situazione per i diritti fondamentali sono state ampiamente analizzate nel corso del lavoro, e già menzionate in più parti di queste conclusioni. La profilazione algoritmica premette di influenzare preferenze, scelte e comportamenti anche in ambiti considerati controversi, ed estremamente rilevanti per la definizione dell'identità di ogni individuo (la polarizzazione politica generata dai *social network* è il primo, ovvio esempio della circostanza). Il controllo di una parte sempre più rilevante del discorso pubblico finisce, di fatto, per essere esclusiva di censori privati senza precedenti, che filtrano, con l'uso massiccio di sistemi automatizzati basati sull'apprendimento automatico, i contenuti diffusi sulle reti sociali di cui sono proprietari. L'analisi dei dati a supporto di meccanismi decisionali può portare al rischio di normalizzare ed istituzionalizzare discriminazioni ingiustificate, e l'uso sempre maggiore di algoritmi per valutazioni e selezioni di ogni genere erode il margine di riservatezza in passato garantito alla persona, rendendo i dettagli più spiacevoli della sua esistenza accessibili con inedita facilità.

È noto che il compito del diritto costituzionale è, prima di tutto, limitare il potere, sottoponendolo alla forza del diritto al fine di sottrarre l'individuo all'arbitrio dei suoi governanti. Le prime costituzioni liberali, a cavallo tra XVIII e XIX secolo, sorgono con questa precipua finalità, da raggiungere, secondo l'arcinota formula dell'art. 16 della *Declaration* del 1789, attraverso due strumenti principali: la separazione dei poteri e la garanzia dei diritti. Solo la creazione di centri di potere complessi e distinti – e dunque in grado di controllarsi e arginarsi a vicenda – può allontanare il rischio di una rapida deriva tirannica, e la garanzia di alcune prerogative inviolabili dell'individuo esplicita i limiti alla potestà dell'autorità sulla persona. Com'è noto, l'impianto

liberale non fu sufficiente a scongiurare l'avvento dei totalitarismi europei del XX secolo, e oggi alle garanzie enunciate più di due secoli fa si aggiungono la rigidità delle carte fondamentali e la presenza, in varie forme, di un controllo di costituzionalità della legge. I diritti individuali, inoltre, sono stati protagonisti di profonde evoluzioni e di aggiornamenti dell'originario, ristretto catalogo liberale, e una delle tesi principali di questo lavoro è che un fenomeno analogo stia accadendo anche in conseguenza dell'avvento delle tecnologie intelligenti.

Tuttavia, l'impianto tradizionale, di cui i testi costituzionali delle principali democrazie contemporanee offrono l'esempio, è volto a limitare lo spazio di azione dei poteri pubblici, percepiti come gli unici in grado di minacciare concretamente la libertà individuale, e, dunque, imbrigliati in un complesso sistema di pesi e contrappesi. Le ragioni sono prima di tutto pratiche: al momento dell'elaborazione di tali costituzioni, le autorità statali sembravano le uniche dotate dei mezzi necessari a convertirsi in un serio rischio per i diritti individuali. In alcuni casi, poi, vi sono evidenti ragioni storiche: diverse carte fondamentali, compresa quella italiana, vedono la propria genesi nella caduta di regimi lunghi e sanguinosi, dei cui abusi i costituenti avevano ben vivo il ricordo. Il problema attuale, invece, è di natura diversa: i diritti fondamentali non sono messi in discussione dall'azione dei poteri pubblici, ma dalla condotta di giganteschi operatori privati, le cui possibilità economiche superano, in diversi casi, quelle di stati sovrani, e sui quali l'esercizio di poteri coercitivi pare molto complesso, in ragione dell'immaterialità dei loro servizi e della radicale novità di questi ultimi. Sono le stesse società che gestiscono i *social network*, di fatto, a dettare molte delle regole che disciplinano le attività al loro interno degli utenti, e ogni disciplina da parte delle autorità pubbliche che riguardi queste ultime si espone a seri problemi di effettività. Il paradigma, dunque, sembra rovesciarsi rispetto all'impostazione tradizionale: i diritti fondamentali sono chiamati in causa da nuovi poteri privati, e i poteri pubblici paiono sprovvisti degli strumenti per reagire, posto che il quadro giuridico esistente si rivela insufficiente.

Far fronte alla situazione è, probabilmente, la principale sfida del costituzionalismo contemporaneo, chiamato a riscoprire la sua originaria natura di limite al potere e adattarla al mutamento di scenario causato dalla rivoluzione digitale e acuito dalle possibilità dischiuse dall'avvento delle tecnologie intelligenti. In sintesi il diritto costituzionale – è questa è un'altra delle tesi essenziali cui è giunta questa ricerca – è chiamato all'elaborazione di nuove strategie (in primo luogo, nell'ottica qui adottata, nuovi diritti) per contrastare l'ascesa di tali inediti, e sempre più potenti, operatori privati, al fine di conservare la sua funzione di garanzia della sfera individuale.

Società algoritmica e società antropocentrica: alcune considerazioni metagiuridiche

Gran parte delle considerazioni finora svolte in queste conclusioni mettono in luce le possibili conseguenze derivanti da funzionamenti dell'intelligenza artificiale non desiderati. I possibili esiti discriminatori di alcuni utilizzi dei sistemi avanzati nell'attività decisionale, ad esempio, di certo non corrispondono ai risultati auspicati al momento del loro sviluppo. Allo stesso modo, ben pochi sviluppatori hanno l'obiettivo di costruire tecnologie in grado di sfuggire totalmente al controllo degli esseri umani, perché radicalmente non interpretabili e, dunque, imprevedibili. Le conseguenze in materia di radicalizzazione dell'opinione pubblica e condizionamento dell'individuo dell'utilizzo di algoritmi di profilazione con finalità di *nudging* e propaganda non sono, in larga parte, desiderate nemmeno da parte delle piattaforme che di tale utilizzo sono le principali responsabili.

Preme ora riportare anche in questa sede alcune considerazioni di carattere metagiuridico svolte in coda alla seconda parte dell'opera, nell'esaurire l'approfondimento delle ripercussioni dell'avvento dell'intelligenza artificiale sull'effettività del principio di eguaglianza. I possibili effetti delle tecnologie avanzate sulla vita degli esseri umani, infatti, non sono limitati all'eventualità di loro malfunzionamenti. In particolare, il crescente coinvolgimento in procedimenti in senso lato decisionali o valutativi di sistemi di apprendimento automatico applicati all'analisi dei dati sembra in grado di mettere in discussione, da vari punti di vista, le basi dell'intera organizzazione sociale. Si è fatto l'esempio di un'ipotetica società futura, in cui gli algoritmi abbiano lentamente soppiantato l'essere umano in gran parte delle decisioni complesse, in ragione della precisione garantita dall'analisi di una mole di informazioni e precedenti sempre maggiore. Ciascuno di noi svolgerebbe il lavoro più adatto alle proprie competenze e capacità, guadagnando un salario esattamente corrispondente al valore di tali mansioni; frequenterebbe i locali, gli amici e i partner sentimentali a lui o lei più consoni, grazie alle determinazioni di sistemi dei quali non comprenderebbe appieno il funzionamento, ma che si rivelerebbero tremendamente efficaci; adotterebbe stili di vita, passioni, abitudini verso le quali sarebbe impercettibilmente e continuamente indirizzato dalle applicazioni tecnologiche da cui sarebbe circondato, delle quali, forse, nemmeno percepirebbe più l'esistenza. Si tratta di uno scenario che deve suscitare riflessioni più ampie di quelle riguardanti gli eventuali errori e risultati imprevisti – rari, ma non eliminabili – delle tecnologie in esame. Infatti, riprendendo alcune delle considerazioni svolte nel corso del lavoro, cosa ne sarebbe, in una società di questo tipo, dell'ambizione a migliorarsi, posto che gran parte delle nostre qualità e dei nostri limiti troverebbe plausibilmente una spiegazione – statistica, ma molto solida – nell'appartenenza a un gruppo di individui con determinate caratteristiche simili? Come verrebbe concepita la diseguaglianza, in un mondo in cui gran parte delle differenze economiche e sociali potrebbero trovare la loro giustificazione nelle valutazioni di un algoritmo? Che accorgimenti verrebbero

predisposti qualora la disponibilità di moli di dati sempre maggiori finisse per svelare la correlazione tra determinate caratteristiche (come l'appartenenza etnica o di genere) e comportamenti considerati spiacevoli o apertamente illegali? L'ordinamento giuridico reagirebbe con misure compensative – in primo luogo in campo educativo – o assumendo un volto meno garantista verso i membri di alcune minoranze?

Gli esempi potrebbero continuare. Il punto, in estrema sintesi, è che la diffusione dell'intelligenza artificiale apre la possibilità di una società *data-driven*, in cui ogni decisione è basata su una mole di informazioni oggi impensabile e, dunque, risulta più precisa. Emerge, però, un dato fondamentale: la consapevolezza della fallibilità delle valutazioni e decisioni umane ha contribuito, nel corso dei secoli, a definire le modalità di interazione tra gli individui che compongono le nostre società. Essa è alla base di alcune delle virtù che, non a caso, siamo abituati a contraddistinguere proprio con la parola "umanità": la comprensione reciproca, il perdono, la solidarietà verso gli errori altrui e la volontà di risolverli. Inoltre, proprio la difficoltà a dare una giustificazione alle differenze, in primo luogo di reddito e condizioni di vita, tra individui, etnie e classi sociali, è stata il motore delle istanze egualitarie che hanno caratterizzato le società democratiche almeno a partire dalla rivoluzione francese. La società algoritmica in esame è una società che potrebbe essere molto meno incline a questi valori, fornendo una spiegazione tecnologica, più facile da considerare "giusta" e razionale, alle differenze. Si tratta di un'evoluzione possibile, che starà in primo luogo al diritto affrontare, al fine di indirizzare lo sviluppo tecnologico in direzione antropocentrica e salvaguardare le fondamentali prerogative appena menzionate. Valori la cui messa in discussione, paradossalmente, sembrerebbe poter derivare non da errori o anomalie di funzionamento di alcuni sistemi avanzati, ma, all'opposto, proprio dal loro corretto funzionamento.

BIBLIOGRAFIA

A.A.V.V., *A programme for advanced information technology. The Report of the Alvey Committee*, Londra, 1982

A.A.V.V., *AI4Peoples 7 AI Global Frameworks*, 2018

A.A.V.V., *Il diritto comparato dell'intelligenza artificiale*, in *Diritto Pubblico Comparato ed Europeo*, numero monografico 1, 2022

A.A.V.V., *Open letter to the European Commission – Artificial Intelligence and Robotics*, 2018

A.A.V.V., *The reality club: one half of a Manifesto*, in *Edge.org*, 11 ottobre 2000

ABALDO G., *Una prospettiva di regolamentazione degli ISP attraverso il Digital Service Act*, in *MediaLaws*, 3 febbaio 2022

ABBAS A., KHALID J., MUBARAK S., JAVED H., *Analyzing the reliability of human social scoring system (HSSS) & its determinants*, in *Journal of marketing and information systems*, 4, 1, p. 33-42

ABDOU H.A., POINTON J., *Credit scoring, statistical techniques and evaluation criteria: a review of the literature,* in *Intelligent systems in accounting, finance and management*, 18, 2-3, p. 59-88

ABRAMS F., *The soul of the first amendment*, New Haven, 2017

ACCESS NOW, *Human rights in the age of artificial intelligence-Report*, 2018

ACKERMAN B., *Before the nextattack. Preservingcivilliberties in an age of terrorism*, New Haven (US), 2007

ACKERMAN B., *The emergencyconstitution*, in *The Yale Law Journal*, 113, 5, 2004, p. 1029-1091

ADADI A., BERRADA M., *Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)*, in *IEEE Access*, 6, 2018, 5244-5245

ADAMS M., *Over-reliance on technology may be at the expense of care*, in *Nursing standard*, 29,6, 2014, p. 32-33

ADETUNJI J., *Tech giants need to take more responsibility for the advertising that makes them billion*, The Conversation, 7 dicembre 2018

ADIWARDANA D. ET AL., *Towards a Human-like Open-Domain Chatbot*, 2020, arXiv:2001.09977

AGGARWAL C.C., *Neural networks and deep learning – a textbook*, Berlino, 2018

AGGARWAL P., PAPAY F.A., *Artificial intelligence image recognition of melanoma and basalcell carcinoma in racially diverse populations*, in *Journal of Dermatological Treatment*, 33, 4, 2022, p. 2257-2262

AGRÒ A.S., ROMAGNOLI U., *Commento all'art. 3 cost.*, in BRANCA G. (A CURA DI), *Commentario della Costituzione italiana*, Bologna, 1976

AHMED I., JEON G., PICCIALLI F., *From artificial intelligence to explainable artificial intelligence in Industry 4.0: a survey on what, how, and where*, in *IEEE Transactions on Industrial Informatics*, 18, 8, p. 5031-5042, 2022

AHN M.J., CHEN Y.C., *Artificial intelligence in government: potentials, challenges, and the future*, in *The 21st annual international conference on digital government research*, Seoul, 15-19 giugno 2020, p. 243 ss

AI FOR GOOD GLOBAL SUMMIT 2017, *Report*, Ginevra, 7-9 giugno 2017

AI Strategy 2019 – Ai for everyone: People, Industries Regions and Governments, 11 giugno 2021

ALBERTINI P., *Article 16*, in CONAC G., DEBENE M., TEBOUL G., *La déclaration des droits de l'homme et du citoyen de 1789, Histoire analyse et commentaires*, Parigi, 1993, p. 331 ss.

ALEMANNO A., SIBONY A.L.(a cura di), *Nudge and the law: a europeanperspective*, Oxford, 2015

ALEXANDER S., *They decide wholives, whodies. Medicalmiracol and a moral burden*, in *Life*, 9, 1972, p. 102 ss.

ALEXI R., *Teoria dell'argomentazione giuridica. La teoria del discorso razionale come teoria della motivazione giuridica*, Milano, 1988

ALEXY R., *Teoria dei diritti fondamentali* (1994), Bologna, 2012

ALFIERI C., EGROT M., DESCLAUX A., SAMS K., *RecognisingItaly'smistakes in the public healthresponse to COVID-19*, in *Lancet*, 399, 10322, 2022, p. 357 ss.

ALGHATRIF M.,LINDSAY J., *A brief review: history to understand fundamentals of electrocardiography*, in *Journal of Community Hospital Internal Medicine Perspectives*, 2, 1, 2012, p. 14383 ss.

ALICIOGLU G., SUN B., *A survey of visual analytics for Explainable Artificial Intelligence methods*, in *Computers & Graphics*, 102, 2022, p. 502 ss.

ALIKHADEMI K., DROBINA E., PRIOLEAU D., *A review of predictivepolicing from the perspective of fairness*, in *Artificial Intelligence and Law*, 30, 2022, p. 1–17

ALKURDI D.A., ILYAS M., ILYAS A., JAMIL A.,*Cancer detection using deep learning techniques* in *Evolutionary Intelligence*, 2021, doi:10.1007/s12065-021-00635-5

ALPA G., RESTA G., *Le persone e la famiglia. 1. Le persone fisiche e i diritti della personalità*, in SACCO R. (diretto da), *Trattato di diritto civile*, Torino, 2019, p. 145 ss.

AlphaGo ha vinto: la macchina ha battuto l'uomo 4-1, La Repubblica, 15 marzo 2016

ALÙ A., *Oversight board di Facebook alla prima prova: così si disvela il suo ruolo*, Agenda digitale, 1 febbraio 2021

AMATO G., *Art. 13* e *Art. 14* in BRANCA G. (A CURA DI), *Commentario della Costituzione*, II, Bologna-Roma, 1977

AMATO S., *Biodiritto 4.0. Intelligenza artificiale e nuove tecnologie*, Torino, 2020

AMIRANTE D., *La reformette dell'ambiente in Italia e le ambizioni del costituzionalismo ambientale*, in *Diritto pubblico comparato ed europeo*, 2, 2022, p. 5 ss.

AMMANNATI L., GRECO G.L., *Il credit scoring alla prova dell'intelligenza artificiale*, in RUFFOLO U. (a cura di), *XXXVI lezioni di diritto dell'intelligenza artificiale*, Torino, 2021, p. 373 ss.

AMOROSO D., *Jus in bello and jus ad bellum arguments against autonomy in weapons systems: a re-appraisal*, in *Questions on International Law,* Zoom in 43, 2017, p. 5 ss.

ANAND R.S. ET AL., *Predicting Mortality in Diabetic ICU Patients Using Machine Learning and Severity Indices*, in *Proceeding from AMIA Summits on Translational Science Proceedings*, 2018, p. 310 ss.

ANDERSON E.S., *Integration, affirmative action and strict scrutiny*, in *New York University Law Review*, 77, 5, 2002, p. 1195 ss.

ANDREWS L., *Public administration, public leadership and the construction of public value in the age of the algorithm and 'big data'*, in *Public Administration*, 97, 2, 2019

ANGELOV P.P., SOARES E.A., JIANG R., ARNOLD N.I., ATKINSON P.K., *Explainable artificial intelligence: an analytical review*, in *WIREs Data Mining & Knowledge Discovery*, 11, 5, 2021

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 maggio 2016

ANTANI R., *The resistance of memory: could the European Union's right to be forgotten exists in the United States?*, in *Berkeley Technology Law Journal*, 30, 385, 2015, p. 1173-1210 ss.

ANTON A.I., EARP J.B., HE Q., STUFFLEBAUM W., BOLCHINI D., JENSEN C., *Financial Privacy Policies and the Need for Standardization*, in *IEEE Security & Privacy*, 36, 2004, p. 42-44

ANTONIOU A., DOSSENA G., MACMILLAN J., HAMBLIN S., CLIFTON D., PETRONE P., *Assessing the risk of re-identification arising from an attack on anonymized data*, 2022, arXiv:2203.16921

APOLLONIO RODIO, *Argonautiche*, IV, 1638-1693

ARASTEH H. ET AL., *Iot-based smart cities: A survey*, in 2016 *IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC)*, 2016, doi: 10.1109/EEEIC.2016.7555867

ARAUJO T., HELBERGER N., KRUIKEMEIER S., DE VREESE C.H., *In AI We Trust? Perceptions about Automated Decision-Making by Artificial Intelligence*, in *AI& SOCIETY*, 35, 3, 2020, p. 611–623

ARCESATI R., *E-government and Covid-19: digital china goes global*, in *MERICS*, 17 marzo 2022

ARCOLEZI H.H., COUCHOT J.F., CERNA S., GUYEUX C., ROYER G., BOUNA B.A. ET AL., *Forecasting the number of firefighter interventions per region with local-differential-privacy-based data*, in *Computers & Security*, 96, 2020, p. 101888 ss.

ARDUINI S., *La "scatola nera" della decisione giudiziaria: tra giudizio umano e giudizio algoritmico*, in *BioLaw Journal – Rivista di BioDiritto*, 2, 2021, p. 463 ss.

ARENSI P., *"I giorni più bui del Covid in ospedale? E' stato come finire per caso in guerra"*, Il Giorno, 31 maggio 2022

ARISTOTELE, *Organon*, a cura di G. COLLI, Torino, 1955.

ARNOBIO, *Adversusnationes*, VI, 22.

ARNOLD T., SCHEUTZ M., *The "big red button" istoo late: an alternative model for the ethicalevaluation of AI systems*, in *Ethics and Information Technology*, 20, 1, 2018

ARNOLD V., COLLIER P., LEECH P.S., SUTTON S.G., *Impact of intelligent decision aids on expert and novice decision-makers' judgments*, in *Accounting & Finance*, 44, 1, 2004, p. 1– 26

ARNOLD V., SUTTON S.G., *The theory of technology dominance: Understanding the impact of intelligent decision aids on decision maker's judgments*, in *Advances in Accounting Behavioral Research*, 1, 3, 1998 p. 175– 194

ARTHUR B., *The second economy*, in *McKinsey Quarterly*, Oct. 2011

ARTICLE 29 DATA PROTECTION WORKING PARTY, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, 6 febbraio 2018

ASARO P., *On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making*, in *International Review of the Red Cross*, 2012, 94, p. 687-709

ASHRAFIAN H., DARZI D., *Transforming health policy through machine learning*, in *PLOS Medicine*, 15, 11, 2018

ASHTON K., *That "Internet of Things" thing*, in *RFID – Journal*, 22 giugno 2009

ASIA PACIFIC FOUNDATION OF CANADA, *Artificial intelligence policies in East Asia: an overview from the Canadian Perspective*, 2019

ASILOMAR AI PRINCIPLES 2017, https://futureoflife.org/ai-principles/

ASIMOV I., *I, Robot*, New York, 1950

ASIMOV I., *Robots and Empire*, New York, 1985

AUBY J.B., DE GREGORIO V., *Le smart cities in Francia. Istituzioni del federalismo*, in *Rivista di studigiuridici e politici*, 4, 2015, p.975-993

AUGUSTINE R., *A critique on content moderation on Facebook. A study based on "stop the steel" conspiracycampaign*, in *IJRCSS*, 21, 2021

AULETTA T., *Riservatezza e tutela della personalità*, Milano, 1978

AUTOR D.H., *Why are there still so many jobs? The history and future of workplace automation*, in *Journal of Economic Perspectives*, 29, 2015, p. 3 ss.

AVANZINI G., *Decisioni amministrative e algoritmi informativi. Predeterminazione, analisi predittiva e nuove forme di intellegibilità*, Napoli, 2019

AZZENA L.M., *L'algoritmo nella formazione della decisione amministrativa: l'esperienza italiana*, in *Revista Brasileira de EstudosPolíticos*, 123, 2021, p. 503-537

BACLIC O., TUNIS M., YOUNG K., DOAN C., SWERDFEGER H., SCHONFELD J., *Challenges and opportunities for public health made possible by advances in natural language processing*, in *Canada Communication Disease Report*, 4, 46, 6, p. 161-168

BACONE F., *La nuova Atlantide*, a cura diGUGLIELMONI P., Milano, 1997

BALDASSARE A., *Privacy e costituzione*, Roma, 1974

BALDASSARRE A., *Diritti della persona e valori costituzionali*, Torino, 1997

BALDUCCI ROMANO F., *La Corte di giustizia "resetta" ildirittoall'oblio*, in *Federalismi.it*, 3, 2020, p. 31-46

BALKIN J.M., *Free speech in the algorithmic society: big data, private governance, and new school speech regulation*, in *U.C. Davis Law Review*, 51, 3, 2018, p. 1149-1210

BANCO M., BRILL E., *Scaling to Very Very Large Corpora for Natural Language Disambiguation*, in *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, 2001, p. 26-33

BARBERA A., *Art. 2*, in BRANCA G. (A CURA DI), *Commentario della Costituzione*, Bologna-Roma, 1975, p. 50 ss.

BARBERA M., AIMO M., *Il nuovo diritto antidiscriminatorio: il quadro comunitario e nazionale*, Milano, 2007

BARBERA M., GUARISO A., *La tutela antidiscriminatoria. Fonti strumenti interpreti*, Torino, 2020

BARILE P., *Diritti dell'uomo e libertà fondamentali*, Bologna, 1984

BARILE P., *Libertà di manifestazione del pensiero*, Milano, 1975

BARKAT S.A., BUSIOC M., *Human–AI Interactions in Public Sector Decision Making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice*, in *Journal of public administration research and theory*, 2022, https://doi.org/10.1093/jopart/muac007

BAROCAS S., SELBST A.D., *Big Data's Disparate Impact*, in *California Law Review*, 104, 2016, p. 671-732

BARRAT J., *Our final invention: artificial intelligence and the end of the human era*, New York, 2013.

BARRATT S., *Internet: Neural introspection for interpretable deep learning, in Proceedings of the Symposium on Interpretable Machine Learning*, Long Beach, CA, USA, 7 December 2017, p. 47–53

BARREDO ARRIETA A., DÍAZ-RODRÍGUEZ N., DEL SER J., BENNETOT A., TABIK S., BARBADO A. ET AL., *Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI*, in *Information Fusion*, 58, 2020, p. 82–115

BARRETT J.T., BARRETT J., *A history of alternative dispute resolution*, San Francisco, 2004

BARRETT P., *Who moderates the social media giants?*, Report - NYU Stern, 2020

BARTOLE S., CONFORTI B., RAIMONDI G., *Commentario alla Convenzione europea dei diritti dell'uomo e delle libertà fondamentali*, Padova, 2001

BARTOLE S., DE SENA P., ZAGREBELSKY V. (A CURA DI), *Commentario breve alla Convenzione europea dei diritti dell'uomo*, Padova, 2012

BARTOLI R., *Il problema della causalità penale. Dai modelli unitari al modello differenziato*, Torino, 2010

BARTOLUCCI L., *Le generazioni future (con la tutela dell'ambiente) entrano "espressamente" in Costituzione*, in *Forum di Quaderni Costituzionali*, 2, 2022, p. 20 ss.

BASILE F., *Diritto penale e intelligenza artificiale*, in *Giurisprudenza italiana*, 172, 2019, p. 67-74

BASILE F., *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, in BASILE F., CATERINI M., ROMANO S. (A CURA DI), *Il sistema penale ai confini delle hard sciences. Percorsi epistemologici tra neuroscienze e intelligenza artificiale*, Ospedaletto (PI), 2021

BASILE F., *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto penale e uomo*, 10, 2019, p. 1-33

BASSINI M., LIGUORI L., POLLICINO O., *Sistemi di intelligenza artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, in PIZZETTI F.(A CURA DI),*Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018

BATHAEE Y., *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, in *Harvard Journal of Law & Technology*, 31, 2, 2018, p. 889-938

BAVETTA G., *Identità (diritto alla)* in *Enciclopedia del diritto*, XIX, 1970, p. 953 ss.

BEAUSOLEIL L., *Is trolling Trump a right or a privilege? The erroneous finding in Knight First Amendment Institute at Columbia University v. Trump*, in *Boston College Law Review*, 60, 9, 2019, p. 31 ss.

BEIL M., PROFT M., VAN HEERDEN D., SVIRI S., VAN HEERDEN P., *Ethical considerations about artificial intelligence for prognostication in intensive care*, in *Intensive Care Medicine Experimental*, 7, 1, 2019

BELLE V., PAPANTONIS I., *Principles and practice of explainable machine learning*, 2020, https://arxiv.org/abs/2009.11698 (2 agosto 2022)

BENANTI P., *Oracoli. Tra algoretica e algocrazia*, Roma, 2018

BENBYA H., DAVENPORT T.H., PACHIDI S., *Artificial Intelligence in Organizations: Current State and Future Opportunities*, in *MIS Quarterly Executive*, 19, 4, 4, 2020

BENESTY M., *L'impartialité de certainsjuges mise à mal par l'intelligence artificielle*, in *Village de la justice*, 25 marzo 2016, https://bit.ly/3qSaYuo (20 agosto 2022)

BENJAMIN R., *Assessing Risk, Automating Racism*, in *Science*, 366, 6464, p. 421–422

BENSINGER G., *How big tech companies responded to the storming of the Capitol*, The New York Times, 11 gennaio 2021

BENSINGER G., *Now social media grows a conscience? Facebook and Twitter are taking action. It's too little, too late*, The New York Times, 13 gennaio 2021

BERGER R., *Ninth Amendment*, in *Cornell Law Review*, 1, 1980, p. 1-26

BERGHEL H., *Malice Domestic: The Cambridge Analytica Dystopia,* in *Computer*, 51, 5, 2018, p. 84–89

BERNARDINI M.G., *Una questione di interpretazione? Note critiche su Raccomandazioni SIAARTI, discriminazione in base all'età ed emergenza sanitaria*, in *BioLaw Journal - Rivista di BioDiritto*, 3, 2020

BERWICK A., *Venezuela is rolling out a new ID card manufactured in China that can track, reward and punish citizens*, in *Business Insider*, 18 novembre 2018, https://bit.ly/3pNC5H0 (8 febbraio 2022)

BESSONE M., *Principi della tradizione e nuove direttive in materia di diritto all'immagine*, in *Foro italiano*, IV, 1974, p. 182-184

BETSUM,,*Poteri pubblici e poteri privati nel mondo digitale*, in *Rivista "Gruppo di Pisa"*, 2, 2021, p. 166-191

BEVERE A., CERRI A., *Il diritto di informazione e i diritti della persona*, Milano, 1995, p. 154 ss.

BHANDARI M., ZEFFIRO T., REDDIBOINA M., *Artificial intelligence and robotic surgery: current perspective and future directions*, in *Current Opinion in Urology*, 30, 1, 2020, p. 48-54

BHARGAVA V.R., VELASQUEZ M., *Ethics of the attention economy: the problem of social media addiction*, in *Business Ethics Quarterly*, 31, 3, 2021, p. 321-359

BHATNAGAR S. ET AL., *Mapping intelligence: requirements and possibilities*, in V.C. MULLER (ED.), *Philosophy and Theory of Artificial Intelligence*, Berlino, 2017

BHATTACHARYYA A., SHEIKHALISHAHI S., DUGAR S., KRISHNAN S., DUGGAL A., OSMANI V., *Predicting delirium risk for the following 24 hours in critically ill patients using deep learning*, in *Journal of Critical Care Medicine*, 48, 1, 2020, p. 182 ss.

Bias (n.), in *Online Etymology Dictionary*, https://www.etymonline.com/word/bias

BICKHARD M.H., TERVEEN L., *Foundational issues in artificial intelligence and cognitive science: impasse and solution*, Amsterdam, 1995

BIFULCO R., CARTABIA M., CELOTTO A. (A CURA DI), *L'Europa dei diritti*, Bologna, 2001

BIFULCO R., *Diritto e generazioni future. Problemi giuridici della responsabilità intergenerazionale*, Milano, 2008

Big tech hasoutgrownthisplanet, The New York Times, 29 luglio 2021

BIGMAN Y., GRAY K., WAYTZ A., ARNESTAD M., WILSON D., *Algorithmic discrimination causes less moral outrage than human discrimination*, PsyArXiv, 2020, doi:10.31234/osf.io/m3nrp

BILANCIA P., DE MARCO E., PIZZETTI F.G., *"Nuovi diritti" e "tutela multilivello dei diritti",* in BILANCIA P., DE MARCO E., PIZZETTI F.G. (a cura di), *L'ordinamento della Repubblica: le istituzioni e la società*, Padova, 2021, p. 517 ss.

BILOTTA F., ZILLI A., *Codice di diritto antidiscriminatorio*. Ospedaletto (PI), 2019

BIN R., *Diritti e argomenti: il bilanciamento degli interessi nella giurisprudenza costituzionale*, Milano, 1992

BIN R., *Nuovi diritti e vecchie questioni*, in A.A.V.V., *Studi in onore di Luigi Costato*, III, Napoli, 2014, p. 75-84

BIN R., *Critica della teoria dei diritti*, Milano, 2018

BIN R., CHIARELLA P., *Critica della teoria dei diritti. Conversazione con Roberto Bin*, in *Ordines – per un sapere interdisciplinare nelle istituzioni europee*, 2, 2018, p. 327 ss.

BISCARETTI DI RUFFIA P., *Uguaglianza (principio di)*, in *Novissimo digesto italiano*, XIX, Torino, 1982, p. 1088 ss.

BISCHOFF P., *Internet censorship 2022: a global map of internet restrictions*, Comparitech – Report, 25 gennaio 2022

BIZIMUNGU J., *Babyl's chatbot to enhance digital healthcare platform*, The New Times, 11 gennaio 2018

BLOCK H.D., *A review of "Perceptrons: an introduction to computational geometry"*, in *Information and control*, 17, 1970, p. 501-522

BLODGETT S. L., O'CONNOR B., *Racial disparity in Natural Language Processing: a case-study of social media African-American English*, Proceedings of the Fairness, Accountability, and Transparency in Machine Learning Conference, 2017, https://arxiv.org/pdf/1707.00061.pdf.

BOBBIO N., *L'età dei diritti*, Torino, 1990

BÖHME R., KÖPSELL S., *Trained to accept? A field experiment on consent dialogs*, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010, p. 2403–2406

BONAVIDES P., *Curso de direitoconstitucional,* XXVIII ed., Malheiros (Brasile), 2013, p. 590 ss.

BOND R.M., FARISS C.J., JONES J.J., KRAMER A.D.I., MARLOW C., SETTLE J.E., FOWLER J.H., *A 61-million-person experiment in social influence and political mobilization*, in *Nature*, 489, 2012, p. 295-298

BONETTI P., *Terrorismo, emergenza e costituzioni democratiche*, Bologna, 2006

BONEZZI A., OSTINELLI M., *Can algorithms legitimize discrimination?*, in *Journal of Experimental Psychology: Applied, 27*, 2, p. 447–459

BOOLE G., *An Investigation of the Laws of Thought*, Cork, 1854

BOOLE G., *The Mathematical Analysis of Logic*, Cork, 1847

BORELLI S., RANIERI M., *La discriminazione nel lavoro autonomo. Riflessioni a partiredall'algoritmo Frank*, in *Labour& Law Issues*, 2021, http://labourlaw.unibo.it/article/view/13169

BORZAGA M., *Le ripercussioni del progressotecnologico e dell'IntelligenzaArtificiale sui rapporti di lavoro in Italia*, in *DPCE online*, 1, 2022, p. 393-403

BORZAGA M., MAZZETTI M., *Discriminazioni algoritmiche e tutela dei lavoratori: riflessioni a partire dall'Ordinanza del Tribunale di Bologna del 31 dicembre 2020*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2022, p. 225-250

BOSTROM N., *Superintelligence: paths, dangers, strategies*, Oxford, 2014

BOTSMAN R., *Who can you trust? How technology brought us together and why it might drive us apart*, New York, 2017

BOTTA E., STANZIONE I., *Una ricerca esplorativa sulla condizione degli studenti stranieri all'ultimo anno del percorso di studi alla Sapienza Università di Roma*, in *Formazione, lavoro, persona*, 36, 2022, p. 150-174

BOTTS T.F., *Forequalsonly: race, equality and the equalprotectionclause*, Lanham, 2018

BOUAZIZ M., *Significations et interprétations de l'article 16 de la Déclaration des droits de l'homme et du citoyen de 1789. Contribution à l'histoire de la notion de constitution,* Parigi, 2019

BOYD D.M., ELLISON N.B., *Social network sites: definition, history, and scholarship*, in *Journal of computer-mediated communication*, 13, 1, 11, 2016

BOYLE P., *Facebook censors support of Kurds*, in *Green Left Weekly*, 1 febbraio 2015

BRACHMAN R.J., *(AA)AI — more than the sum of its parts - 2005 AAAI Presidential Address*, in *AI Magazine*, 27, 4, 2006, p. 19 ss.

BRANTINGHAM P.J., *The Logic of Data Bias and its Impact on Place-Based Predictive Policing*, in *Ohio State Journal of Criminal Law*, 15, 2017, p. 473 ss

BRANTINGHAM P.J., VALASIK M., MOHLER G.O., *Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial*, in *Statistics and Public Policy*, 5, 1, 2018

BRASIL S., PASCOAL C., FRANCISCO R., DOS REIS FERREIRA V.A., VIDEIRA P., VALADÃO G., *Artificial Intelligence (AI) in Rare Diseases: Is the Future Brighter?*, in *Genes*, 10, 12, 2019, p. 978 ss.

BRAUN J.A., EKLUND J.D., *Fake News, Real Money: Ad Tech Platforms, Profit-Driven Hoaxes, and the Business of Journalism*, in *Digital Journalism*, 7, 1, p. 1–21

BRIDE H., DONG J., DONG J.S., HÓU Z., *Towards dependable and explainable machine learning using automated reasoning*, in *Proceedings of the International Conference on Formal Engineering Methods*, Gold Coast, Australia, 12–16 November 2018, p. 412–416

BRINGSJORD S., GOVINDARAJULU N.S., *Artificial Intelligence*, in ZALTA E.N. (a cura di), *The Stanford Encyclopedia of Philosophy*, 2020, https://stanford.io/384A1Bl (26 novembre 2021)

BRISON S.J., GELBER K., *Free speech in the digital age*, Oxford, 2019

BRKAN M., *Artificial Intelligence and Democracy: The Impact of Disinformation, Social Bots and Political Targeting*, in *Delphi - Interdisciplinary Review of Emerging Technologies*, 2, 2019, p. 66-71

BROCKDORFF N., APPLEBY-ARNOLD S., *What consumers think*, EU CONSENT Project, Workpackages 7-8, 2013

BRODIE A., VASDEV N., *The future of robotic surgery*, in *Annals of the Royal College of Surgeons of England*, 100, 7, 2018, p. 4-13.

BRONZINI G., COSIO R. (A CURA DI), *Interpretazione conforme, bilanciamento dei diritti e clausole generali*, Milano, 2017

BROOKS R.A., *Intelligence without reason*, in STEELS L., BROOKS R.A. (a cura di), *The artificial life route to artificial intelligence*, Mahwah (New Jersey), 1995, p. 57

BROWN L.X.Z., RICHARDSON M., SHETTY R., CRAWFORD A., HOAGLAND T., *Challenging the Use of Algorithm-drivenDecision-making in Benefits DeterminationsAffecting People with Disabilities*, Center for Democracy & Technology – Report, Oct. 2020

BROWN D., *Britain First defies ban on Facebook*, in *The Times*, 15 marzo 2018

BROWN N., SANDHOLM T., *Superhuman AI for multoplayer poker*, in *Science*, 365, 6456, p. 885-890

BRUCH E., FEINBERG F., *Decision-Making Processes in Social Contexts*, in *Annual Review of Sociology*, 43, 1, 2017, p. 207–227

BRUNER J.S., *Intention in the structure of action and interaction*, in LIPSITT L.P., ROVEE-COLLIER C.K. (A CURA DI), *Advances in infancy research,* 1, Norwood, 1996, p. 41 ss.

BUCHANAN B.G., SHORTLIFFE E.H., *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project*, Reading (USA), 1994

BUÇINCA Z., LIN P., GAJOS K., GLASSMAN E.L., *Proxy tasks and subjective measures can be misleading in evaluating explainable AI systems*, in *Proceedings of the 25th International Conference on Intelligent User Interfaces - ACM, Cagliari* (Italy), 2020, p. 454-464

BUDZINSKI A.C., *Reforming service of process: an access-to-justice framework*, in *University of Colorado Law Review*, 90, 1, 2019

BUITEN M.C., *Towardsintelligentregulation of artificial intelligence*, in *European Journal of Risk Regulation*, 10, 1, 2019, p. 41-59

BUNT A., LOUNT M., LAUZON C., *Are explanations always important? A study of deployed, low-cost intelligent interactive systems*, in *Proceedings of the 2012 ACM international conference on IUI*, 2012, p. 169 ss.

BURNAP P., WILLIAMS M.L., *Cyber hate speech on Twitter: an application of machine classification and statistical modeling for policy and decision making*, in *Policy & Internet*, 2015, 7, 2, p. 223–242

BURRETT T., *Journalism in Myanmar: Freedom, Facebook and fake news*, in MORRISON J., BIRKS J., BERRY M. (A CURA DI), *The Routledge Companion to Political Journalism*, Londra-New York, 2022

BURT A., *The AI transparency paradox*, in *Harvard Business Review*, 13 dicembre 2019

BUSTAMANTE J., *Hacia la cuartageneración de DerechosHumanos: repensando la condiciónhumana en la sociedad tecnologica*, in *Revista iberoamericana de ciencia, tecnología, sociedad e innovación*, 1, 2001

BUSTAMANTE T., DAHLMAN C., *Argumenttypes and fallacies in legalargumentation*, Cham, 2015

BUTLER S., *Erehwon*, a cura di DRUDI DEMBY L., Milano, 1993

BUTTARELLI G., *A smart approach: counteract the bias in artificial intelligence*, 8 novembre 2016, https://bit.ly/3JxjWUJ

BUTTURINI D., *La tutela dei diritti fondamentali nell'ordinamento costituzionale italiano ed europeo*, Napoli, 2009

CADALANU G., *Coronavirus, le bufale sull'esercitazione Defender Europe e l'invio di soldati USA*, la Repubblica, 13 marzo 2020

CAHN A., ALFED S., BARFORD P., MUTHUKRISHNAN S., *An empirical study of web cookies*, in *WWW '16: Proceedings of the 25th International Conference on World Wide Web*, 2016, p. 891-901

CAIA A., *Art. 22*, in RICCIO G.M., SCORZA G., BELISARIO E. (A CURA DI),*GDPR e normativa privacy. Commentario*, 2018, p. 219 ss.

CALABRESI G., BOBBITT P., *TragicChoices*, New York-Londra, 1978

CALAFÀ L., GOTTARDI D., *Il diritto antidiscriminatorio tra teoria e prassi applicativa*, Roma, 2009

CALAMANDREI P., *Processo e democrazia*, Padova, 1954

CALLAHAN D., *La medicina impossibile. Le utopie e gli errori della medicina moderna*, Milano, 2009

CAMBRIDGE CONSULTANTS, *Report produced on behalf of Ofcom - Use of AI in online content moderation*, 2019

CAMPBELL E.M., SITTIG D.F., GUAPPONE K.P., DYKSTRA R.H., ASH J.S., *Overdependence on technology: An unintended adverse consequence of computerized provider order entry*, in *AMIA Annual Symposium Proceedings*, 2007, p. 94– 98

CAMPIANI M.L., *Giusto processo civile e penale*, Napoli, 2014

CAPES T., COLES P., CONKIE A., GOLIPOUR L. ET AL., *Siri On-Device Deep Learning-Guided Unit Selection Text-to-Speech System*, in *ProceedingsInterspeech*, 2017, p. 4011-4015

CAPLAN J.M., KENNEDY L.W., *Risk Terrain Modeling: Crime Prediction and Risk Reduction*, Berkeley, 2016

CAPLAN R.L., *The History and Meaning of the Ninth Amendment*, in *Virginia Law Review*, 69, 2, 1983, p. 223-268

CARAPELLA FIGLIA G., *Il divieto di discriminazione quale limite all'autonomia contrattuale*, in *Rivista di diritto civile*, 2015, p. 1387-1418

CARAVITA B., *Oltre l'uguaglianza formale. Un'analisi dell'art. 3 c. 2 della Costituzione*, Padova, 1984

CARDARELLI F., ZENO-ZENCOVICH V., *Il diritto delle telecomunicazioni: principi, normativa, giurisprudenza*, Bari, 1997

CARDILLO I., *Disciplina dell'intelligenza artificiale e intelligentizzazione della giustizia*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2022, p. 139-167

CARDONE A., *La tutela multilivello dei diritti fondamentali*, Milano, 2012

CARETTI P., *I diritti fondamentali. Libertà e Diritti sociali*, Torino, 2005

CARETTI P., *La disciplina della radiotelevisione tra diritto interno e diritto comunitario*, in ANGOTTI F., PELOSI G., *Il telefono e dintorni: una selezione di eventi, contributi ed immagini dalle celebrazioni per il bicentenario della nascita di Antonio Meucci*, 2011, p. 125-128

CARETTI P., TARLI BARBERI G., *I diritti fondamentali*, Torino, 2017

CARNEIRO D., NOVAIS P., ANDRADE F., ZELEZNIKOW J., NEVES J., *Online dispute resolution: an artificial intelligence perspective*, in *Artificial Intelligence Review*, 41, 2, 2014

CARNEVALE F.A., *Moral distress in the ICU: it's time to do something about it*, in *Minerva anestesiologica*, 86, 4, 2020

CARRUTHERS P., SMITH P.K. (A CURA DI), *Theories of theories of mind*, Cambridge, 1996

CARTABIA M. (A CURA DI), *I diritti in azione*, Bologna, 2007

CARTABIA M., *I "nuovi" diritti*, in *Stato, chiese e pluralismo confessionale*, 2, 2011

CARTABIA M., VETTOR T. (A CURA DI), *Le ragioni dell'uguaglianza*, Milano, 2009

CARTER E.L., *Argentina'sright to be forgotten*, in *Emory International Law Review*, 27, 2013, p. 23-41

CARUSI D., *Principio di eguaglianza, diritto singolare e privilegio. Rileggendo i saggi di Pietro Rescigno*, Napoli, 1998. VENNEMAN N., *The German Draft Legislation On the Prevention of Discrimination in the Private Sector*, in *GermanLaw Journal*, 3, 3, 2002

CASATI D., PENNISI M., *La corte suprema di Facebook: chi sono le 20 personalità che hanno deciso sul bando di Trump*, Corriere della Sera, 29 maggio 2021

CASELLI M. ET AL., *Stop worrying and love the robot: an activity-based approach to assess the impact of robotization on employment dynamics*, in *GLO Discussion Paper*, Essen, 2021, 802, p. 30 ss.

CASERTA S., *The sociology of the legal profession in the digital age*, in *International Journal of the Legal Profession*, 2021, doi:10.1080/09695958.2021.1920417.

CASETTA E., *Diritti pubblici subiettivi*, in *Enciclopedia del diritto*, Milano, 1964, XII 791 ss.

CASEY B., FARHANGI A., VOGL R., *Rethinking explainable machines: the GDPR's "right to explanation" debate and the rise of algorithmic audits in enterprise*, in *Berkeley Technology Law Journal,* 34, 1, 2019, p. 143 ss.

CASINI A., *Lo Stato nell'era di Google. Frontiere e sfide globali*, Milano, 2020

CASO R., *La società della mercificazione e della sorveglianza: dalla persona ai dati*, Milano, 2021, p. 99-120

CASONATO C., *Introduzione al biodiritto*, Torino, 2012

CASONATO C., *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, in *BioLaw Journal – Rivista di Biodiritto*, Special Issue 2, 2019, p. 713 ss.

CASONATO C., *Intelligenza artificiale e dirittocostituzionale: prime considerazioni*, in *Diritto pubblico comparato ed europeo*, numero speciale, 2019, p. 110 ss.

CASONATO C., *Intelligenza artificiale e giustizia: potenzialità e rischi*, in *DPCE Online*, 44, 3, 2020, p. 379 ss.

CASONATO C., *AI and constitutionalism: the challenges ahead*, in BRAUNSCHWEIG B., GHALLAB M., *Reflections on Artificial Intelligence for Humanity*, Berlino, 2021, p. 127 ss.

CASONATO C., *Giustizia e intelligenza artificiale: considerazioni introduttive*, in *BioLaw Journal - Rivista di BioDiritto*, 2, 2021, p. 359 ss.

CASONATO C., relazione *The Rise of New (and old) Rights in the Age of AI*, presso l'incontro inaugurale del corso *Constitutional Law of Technologies*, Firenze, 6 ottobre 2021

CASONATO C., *L'intelligenza artificiale e il diritto pubblico comparato ed europeo*, in *DPCE Online***,** 51, 1, 2022

CASONATO C., *Intelligenza artificiale e medicina: l'impatto sulla relazione di cura (cenni)*, in U. SALNITRO (a cura di), *SMART – La persona e l'infosfera. Atti del Convegno 30 settembre - 2 ottobre 2021Catania*, Pisa, 2022, p. 107 ss.

CASONATO C., MARCHETTI B., *Prime osservazioni sulla Proposta di Regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *BioLaw Journal-Rivista di BioDiritto,* 3, 2021

CASSESE S., *La disciplina legislativa del procedimento amministrativo. Una analisicomparata*, in *Il ForoItaliano*, 116, 1, 1993, p. 27-34

CASSESE S., *Chi ha paura delle autorità indipendenti?*, in *Mercato Concorrenza Regole*, 3, 1999, p. 471–474, https://doi.org/10.1434/78

CASSESE A., *I diritti umani oggi*, Roma, 2008

CASSESE S., *L'eguaglianza sostanziale nella Costituzione: genesi di una norma rivoluzionaria*, in *Le carte e la storia*, 1, 2017, p. 5-13

CASSETTI L. (A CURA DI), *Diritti, principi e garanzie sotto la lente dei giudici di Strasburgo*, Napoli, 2012

CASTANGIA I., BIAGIONI G., *Il principio di non discriminazione nel diritto dell'Unione europea*, Napoli, 2011

CASTELLS M., *The Rise of the Network Society - The Information Age: Economy, Society and Culture*, I, Oxford, 2000

CASTELVECCHI D., *Can we open the black box of AI?*, in *Nature*, 2016, 538(7623), p. 20 ss.

CECCHERINI E., *La codificazione dei diritti nelle recenti costituzioni*, Milano, 2002, p. 122 ss.

CELLAN-JONES R., *Stephen Hawking warnsartificial intelligence could end mankind*, BBC News – Technology, 2 dicembre 2014

CELOTTO A., *Le declinazioni dell'eguaglianza*, Napoli, 2011

CENDON P., NEGRO A., *Danno biologico e tabelle milanesi*, Milano, 2011

CENTER FOR EUROPEAN POLICY STUDIES, *Report - Clarifying the Costs for the EU'sAI Act*, Bruxelles, 24 settembre 2021, https://bit.ly/3QGqa9o (20 maggio 2022)

CERRI A., *Uguaglianza (principio costituzionale di)*, in *Enciclopedia giuridica Treccani*, Roma, 1988

CERRI A., *Identità personale*, in *Enciclopedia giuridica*, agg. IV, Roma, 1995

CERRI A., *Riservatezza (diritto alla) II – diritto comparato e straniero*, in *Enciclopedia giuridica*, XXVII, 1995.

CHAFEE Z., *Free speech in the United States*, Cambridge (USA), 1941

CHAN Y.C.,BERTINI E.,NONATO L.G., BARR B., SILVA C.T., *Melody: Generating and Visualizing Machine Learning Model Summary to Understand Data and Classifiers Together*, 2007, http://arxiv.org/abs/2007.10614

CHANG C.C., THOMPSON B., WANG H., NESPEREIRA C.G., ELHARIRI E., EL-BENDARY N., VILAS A.F., REDONDO R.P.D., *Machine Learning Based Classification Approach for Predicting Students Performance in Blended Learning*, in GABER T., HASSANIEN A.E., EL-BENDARY N., DEY N. (A CURA DI), *The 1st International Conference on Advanced Intelligent System and Informatics (AISI2015)*, Cham, 2016, p. 47-56

CHARNIAK E., MCDERMOTT D., *Introduction to artificial intelligence*, Boston, 1985

Chartedudroit à l'oublidansles site collaboratifs et lesmoteurs de recherche, 2010, https://www.huntonak.com/files/webupload/PrivacyLaw_Charte_du_Droit.pdf

CHEN H., ENGKVIST O., WANG Y., OLIVECRONA M., BLASCHKE T., *The rise of deep learning in drug discovery*, in *Drug Discovery Today*, 23, 6, 2018, p. 1241-1250

CHEN Y., CHEUNG A.S., *The trasparent self under big data profiling: privacy and chinese legislation on the social credit system*, in *The journal of comparative law*, 12, 2, p. 356-378

CHENG J. ET AL., *COVID-19 mortality prediction in the intensive care unit with deep learning based on longitudinal chest X-rays and clinical data*, in *European Radiology*, 32, 7, 2022

CHENG Y. ET AL., *Reliability Prediction and Safety Evaluation of ATC Automation System*, in *2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, 2020

CHESINI F., *"Terminator scenario"? Intelligenza artificiale nel conflitto armato: "lethalautonomousweapons systems" e le risposte del diritto internazionale umanitario*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2020, p. 441-471

China's Social Credit System in 2021: From fragmentation towards integration, MERICS – Report, 3 marzo 2021

CHITI E., MARCHETTI B., RANGONE N., *L'impiego di sistemi di intelligenza artificiale nelle pubbliche amministrazioni italiane: prove generali*, in *BioLaw Journal - Rivista di BioDiritto*, 2, 2022

CHITI E., MARCHETTI B., RANGONE N., TOGNA T., LIMOSANI A., FREGA G., DERIU P., RAGUCCI G., LO SCHIAVO L., LAZZA L., *Rapporto 1/2021 – L'impiego dell'IA nell'attività di CONSOB, AGCOM e ARERA*, in *BioLaw Journal – Rivista di BioDiritto*, 4, 2021, p. 211 ss.

CHOI Q., *7 revealing ways AIs fail,* in *IEEE Spectrum*, 21 settembre 2021

CHOU Y., MOREIRA C., BRUZA P., OUYANG C., JORGE J., *Counterfactuals and causability in explainable artificial intelligence: theory, algorithms, and applications*, in *Information Fusion*, 81, 2022, p. 59-83

CHOW D.S. ET AL., *Development and externalvalidation of a prognostic tool for COVID-19 criticaldisease*, in *PLoS One*, 15, 12, 2020

CHRONOWSKI N., *Human rights in a multilevel constitutional area*, Parigi, 2018

CHURCHLAND P.M., CHURCHLAND P.S., *Could a machine think?,* in *Scientific American*, 262, 1, 1990, p. 32-39

CITRON D.K., CHESNEY R., *Deep fakes: a looming challenge for privacy, democracy, and national security*, in *California Law Review*, 107, 2019, p. 1753 ss.

CLARICH M., *Autorità Indipendenti: Bilancio e Prospettive Di Un Modello*, Bologna, 2005

CLARK G., *The industrial revolution*, in AGHION P., DURLAUF S.N., *Handbook of economic growth*, II, Amsterdam, 2014, p. 217-262

CLARK R.L., HAMMOND R., SANDLER G., MORRILL M., KHALAF C., *Nudging retirement savings: a field experiment on supplemental plans*, Working Paper 23679 – National Bureau of Economic Research, Cambridge (USA), 2017

CLARKE R., *Asimov's laws of robotics: implications for information technology*, in *Computer*, Dec. 1993, p. 53-61 (pt. I) e Jan. 1994, p. 57-66 (pt. II).

COCCHIARA E., *Procedimento amministrativo e "buon algoritmo"*, in *amministrativ@mente*, 3, 2020, p. 370-385

COGLIANESE C., LEHR D., *Regulating by robot: Administrative Decision Making in the Machine Learning Era*, in *Georgetown Law Journal*, 2017, 105, p. 1147 ss.

COGLIANESE C., LEHR D., *Transparency and AlgorithmicGovernance*, in *AdministrativeLaw Review*, 71, 7, 2019

COHEN A.L., *New guarantees for cryptographiccircuits and data anonymization*, Cambridge, 2019, p. 235 ss.

COHEN J.E., *Law for the platform economy*, in *U.C. D. Law Review*, 51, 2018, p. 133 ss.

COHEN J.P. ET AL., *Predicting COVID-19 Pneumonia Severity on Chest X-ray With Deep Learning*, in*Cureus*,12, 7, 2020

COHEN P.R., *Empirical methods for artificial intelligence*, Cambridge (US), 1995

COLE M.D., ETTELDORF C., CARSTEN U., *Updating the Rules for Online Content Dissemination: Legislative Options of the European Union and the Digital Services ActProposal*, 2022, https://www.nomos-elibrary.de/index.php?doi=10.5771/9783748925934

COLLET B., *Epidémie de grippe: «En dix-neufans de métier, je n'ai jamais vu ça»*, le Monde, 12 gennaio 2017

COMITATO NAZIONALE PER LA BIOETICA, *Covid-19: la decisione clinica in condizioni di carenza di risorse e il criterio del "triage in emergenza pandemica"*, 8 aprile 2020

COMMINS J., *Nurses say distractions cut bedside time by 25%*, in *Health Leader*, 2010

COMMISSIONE EUROPEA, *Tackling online disinformation: a European approach* (COM(2018)236 final)

COMMISIONE EUROPEA, *Intelligenza artificiale per l'Europa* (COM(2018) 237 final)

COMMISIONE EUROPEA, *Coordinated plan on artificiale intelligence* (COM(2018) 795 finalAnnex I)

COMMISIONE EUROPEA, *Libro bianco sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia* (COM(2020) 65 final)

COMMISIONE EUROPEA, *Plasmare il futuro digitale dell'Europa* (COM(2020) 67 final)

COMMISIONE EUROPEA, *Coordinated plan on artificiale intelligence 2021 review* (COM(2021) 205 finalAnnex)

COMMISIONE EUROPEA, *DESI country overview – Italy*, https://digital-strategy.ec.europa.eu/en/policies/desi, 2022

COMMISIONE EUROPEA, *Dichiarazione europea sui diritti e i principi digitali per il decennio digitale*, (COM(2022) 28 final).

COMPAGNUCCI M.C., FENWICK M., HAAPIO H., *Digital technology, future lawyers and the computablecontract designer of Tomorrow*, in COMPAGNUCCI M.C., FENWICK M., HAAPIO H. (A CURA DI), *ResearchHandbook in Contract Design*, Cheltenham-Northampton, 2022, p. 421-445

CONFALONIERI R., COBA L., WAGNER B., BESOLD T.R., *A historical perspective of explainable Artificial Intelligence, in WIREs Data Mining and Knowledge Discovery*, 1, 2021

CONGIU M., *Repubblica Ceca, legge su risarcimento donne Rom sottoposte a sterilizzazione forzata*, il Manifesto, 1 luglio 2021

CONNETT I., *France resistsjudicial AI revolution. France bans predicative analysis of caselaw. Does this protect or impede universal justice?*, in *Above the law*, 10 giugno 2019

CONSIGLIO DI STATO DELLA REPUBBLICA POPOLARE CINESE, doc. n. 35, 8 luglio 2017

CONSIGLIO E., *Che cos'è la discriminazione? Un'introduzione teorica al diritto antidiscriminatorio*, Torino, 2020

COOLEY T.M., *A Treatise on the Law of Torts or the Wrongs Which Arise Independently of Contract*, Chicago, 1880

COOPER C., *Technology and development in the industrial devolution*, Londra, 2005

COOPER J., *Cognitive Dissonance: Where We've Been and Where We're Going*, in *International Review of Social Psychology*, 32, 1, 2019

COPI I.M., COHEN C., RODYCH V., *Introduction to logic (15ª ed.)*, Londra, 2019

CORBYN Z., *Facebook experiment boosts US voter turnout*, in *Nature*, 2012, https://doi.org/10.1038/nature.2012.11401

CORSI S., *Nasce la Corte suprema di Facebook: si chiama Oversight Board*, in *Cyberlaws*, 30 novembre 2020

CORTESE F., BORGONOVO RE D., FLORENZANO D., *Diritti inviolabili, doveri di solidarietà e principio di eguaglianza*, Trento, 2015

CORTEZ P., EMBRECHTS M.J., *Opening black box data mining models using sensitivity analysis*, in *Proceedings of the Symposium on Computational Intelligence and Data Mining (CIDM)*, Paris, France, 11–15 aprile 2011, p. 341–348.

Cose che noi umani. La pandemia che ha sconvolto le nostre vite e resterà per sempre nell'immaginario comune. *Una cronistoria degli eventi che non avremmo mai potuto immaginare*, Lab24 – Il Sole 24 Ore, https://bit.ly/3TcWDF3

COSTANTINI S., *I social sgomberano CasaPound*, in *La Repubblica*, 10 settembre 2019

COUNCIL OF BARS AND LAW SOCIETIES OF EUROPE, *Guide on the use of artificial intelligence-based tools by lawyers and law firms in the EU*, 2022

COUTURAT V.L., *La logique de Leibniz d'après des documents inédits*, Parigi, 1901, p. 81 ss.

CRANOR L.F., *A framework for reasoning about the human in the loop*, in *Proceedings of the 1st Conference on Usability, Psychology, and Security*, USENIX Association, USA, 2008, https://dl.acm.org/doi/10.5555/1387649.1387650

CREEMERS R., *China's social credit system: an evolving practice of control*, 2018, http://dx.doi.org/10.2139/ssrn.3175792

CREMONA E., *Le nuove tecnologie oltre la "grande dicotomia" tra pubblico e privato*, in *Gruppo di Pisa. Dibattito aperto sul Diritto e la Giustizia costituzionale*, Quad. Monografico n. 3, 2, 2021, p. 681 ss.

CREVIER D., *AI: The TumultuosSearch for Artificial Intelligence*, New York, 1993

CRIDDLE C., *Facebook moderator: every day was a nightmare*, BBC News, 12 maggio 2021

CRISAFULLI V., NOCILLA D., *Nazione*, in *Enciclopedia del diritto*, XXVII, 1977, p. 805 ss.

CRISTOFARO M., *Reducing Biases of Decision-Making Processes in Complex Organizations*, in *Management Research Review*, 40, 3, 2017, p. 270–291

CROSS M., *Japan's fifth generation computer project successes and failures*, in *Futures*, 21, 4, 1989, p. 401-403

CULOTTA E., *On the Origin of Religion*, in *Science*, 326, 5954, 2009, p. 784-787

CUNNINGHAM M.L., REGAN M.A., *Driver distraction and inattention in the realm of automated driving*, in *IET Intelligent Transport Systems*, 12, 6, 2018

CUPPINI L., *Covid, il rischio di morte dei pazienti calcolato da un algoritmo*, Corriere della Sera, 5 settembre 2020

CURCIO M., *La dichiarazione dei diritti delle Nazioni Unite*, Milano, 1950

CURINI L., *Da Trump a Q-Anon, se la censura è un boomerang (anche per gli 007)*, in *Formiche.net*, 10 gennaio 2021

CURTIS K., *Wearable tech in healthcare: top devices making a difference*, in *Edumed*, https://bit.ly/3RwWiwv

CURTIS M.K., *No State shall abridge. The XIV Amendment and the Bill of Rights*, Durham (USA), 1986

D. LESLIE, *Understanding artificial intelligence ethics and safety: a guide for the responsible design and implementation of AI systems in the public sector*, The Alan Turing Institute, 2019

D'ADDA A., *La Corte di Cassazione riafferma il proprio orientamento in tema di diritto all'identità personale*, in *Responsabilità civile e previdenza*, 2-3, 1997, p. 474-481

D'ALESSANDRO J., *Nasce la "Corte suprema" di Facebook. Indipendente, giudicherà le scelte del social network*, la Repubblica, 7 maggio 2020

D'ALOIA A., *Eguaglianza sostanziale e diritto diseguale*. *Contributo allo studio delle azioni positive nella prospettiva costituzionale*, Padova, 2002

D'ALOIA A., *Il diritto verso "il mondo nuovo". Le sfide dell'Intelligenza Artificiale*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019

D'ALOIA A. (a cura di), *Intelligenza artificiale e diritto: come regolare un mondo nuovo*, Milano, 2020

D'ALOIA A., *Costituzione ed emergenza. L'esperienza del coronavirus*, in *BioLaw Journal – Rivista di BioDiritto*, Special Issue 1, 2020, p. 7-12

D'ANDREA L., MOSCHELLA G., RUGGERI A., SAITTA A. (A CURA DI), *La Carta dei diritti dell'Unione europea e le altre Carte (ascendenze culturali e mutue implicazioni)*, Torino, 2016

D'ANTENA A., *Tutela dei diritti fondamentali e costituzionalismo multilivello: tra Europa e stati nazionali*, Milano, 2004

D'AVACK L., *CoViD-19: criteri etici*, in *BioLaw Journal - Rivista di BioDiritto*, 1S, 2020

DAL CANTO F., *Lezioni di ordinamento giudiziario*, Torino, 2018

DAN V., PARIS B., DONOVAN J., HAMELEERS M., ROOZENBEEK J., VAN DER LINDEN ET AL. S., *Visual Mis- and Disinformation, Social Media, and Democracy*, in *Journalism & Mass Communication Quarterly*, 98, 3, 2021, p. 641–664

DANAHER J., *Tragic Choices and the Virtue of Techno-Responsibility Gaps*, in *Philosophy & Technology*, 35, 2, 2022

DANKS D., LONDON A.J., *Algorithmic bias in autonomous systems*, in *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*, 17, 2017, p. 4691-4697

DARA R.,*Intermediary Liability in India: Chilling Effects on Free Expression on the Internet*, 2011, http://dx.doi.org/10.2139/ssrn.2038214

DAS A., RAD P., *Opportunities and Challenges in Explainable Artificial Intelligence (XAI)*, https://arxiv.org/abs/2006.11371, 2020

DAVENPORT T., KALAKOTA R., *The potential for artificial intelligence in healthcare*, in *Future Healthcare Journal*, 6, 2, 2019, p. 94-98

DAVID F.N., *Forces of Production: A Social History of Industrial Automation*, New York, 2017

DAVIDSON T., WARMSLEY D., MACY M., WEBER I., *Automated hate speech detection and the problem of offensive language*, 2017, http://arxiv.org/abs/1703.04009

DAVIS A.E., *The Future of Law Firms (and Lawyers) in the Age of Artificial Intelligence*, American Bar Association, 2 ottobre 2020,

DAVIS N., *What is the fourth industrial revolution?*, World Economic Forum, 16 gennaio 2016

DAWS R., *The NHS hopes an AI chatbot will help tackle patient wait times*, in *AI News*, 29 luglio 2022

DE BRUIJN H.,WARNIER M., JANSSEN M., *The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making*, in *Government Information Quarterly*, 39, 2, 2022, p. 101666 ss.

DE CODT J., *Justice et algorithme: danger pour le procèséquitable et la démocratie?*, in *Revuetrimestrielledesdroits de l'homme*, 117, 1, 2019, p. 3-11

DE CUPIS A., *I diritti della personalità*, Milano, 1982

DE FALCO V., *Azione amministrativa e procedimenti nel diritto comparato*, Padova, 2018

DE FELICE D., *Il piano pandemico inadeguato e obsoleto. Ne ho parlato Francesco Zambon*, il Fatto Quotidiano, 7 luglio 2021

DE FINE LICHT K., DE FINE LICHT J., *Artificial Intelligence, Transparency, and Public Decision-Making: Why Explanations Are Key When Trying to Produce Perceived Legitimacy*, in *AI & SOCIETY*, 35, 4, 2020, p. 917–926

DE GREGORIO G., *The Rise of Digital Constitutionalism in the European Union*, in *International Journal of ConstitutionalLaw*, 19, 1, 2021, p. 41–70

DE MIGUEL BERIAIN I., *Does the Use of Risk Assessments in Sentences Respect the Right to Due Process? A Critical Analysis of the Wisconsin v. Loomis Ruling*, in *Law, Probability and Risk*, 17, 1, p. 45–53

DE MINICO G., *Costituzione, emergenza e terrorismo*, Napoli, 2016

DE NARDIS L., HACKL A.M., *Internet governance by social media platforms*, in *Telecommunications Policy*, 39, 9, 2015, p. 761-770

DE OLIVEIRA F.L. ET AL, *Path and future of artificial intelligence in the field of justice: a systematic literature review and a research agenda*, in *SN Social Sciences*, 2, 180, 2022

DE PAOLI S., *Automatic-Play and Player Deskilling*, in *MMORPGs,Game Studies*, 13, 1, 2013

DE PASQUALE P., *Verso una Carta dei diritti digitali (fondamentali) dell'Unione Europea?*, in *Il diritto dell'Unione Europea*, 3, 2022

DE SIMONE E., *Storia economica. Dalla rivoluzione industrial allarivoluzione informatica*, Milano, 2014

DE TULLIO M.F., *Uguaglianza sostanziale e nuove dimensioni della partecipazione politica*, Napoli, 2020

DE VRIES J., *The industrial revolution and the industrious revolution*, in *The journal of economic history*, 54, 2, 1994, p. 249-270

DEAR K., *Artificial Intelligence and Decision-Making*, in *The RUSI Journal*, 164, 5–6, 2018 p. 18–25

Deepfake, in *Techopedia*, https://www.techopedia.com/definition/33835/deepfake

DEIBERT R.J., *The road to digital unfreedom*, in *Journal of Democracy*, 30, 1, 2019, p. 25-39

DEL VICARIO M. ET AL., *The Spreading of Misinformation Online*, in *Proceedings of the National Academy of Sciences*, 113, 3, p. 554–559

DELACROIX S., LAWRENCE N.D., *Bottom-up data trusts: disturbing the "one size fitsall" approach to data governance*, in *International Data Privacy Law*, 9, 4, 2019, p. 236-252

DELLA GIUSTINA C., *Il problema della vulnerabilità nelle Raccomandazioni SIAARTI e nelle linee guida SIAARTI-SIMLA*, in *Stato, Chiese e pluralismo confessionale*, 2021

DELOITTE, *The future of AI in healthcare. How AI will impact patients, clinicians and the pharmaceutical industry*, 2019

DELOITTE,*Bringingtransparency and ethics in AI*, 2019

DENG L., YU D., *Deep Learning: Methods and Applications*, in *Foundations and Trends in Signal Processing*, 7, 3-4, p. 198-205

DENHAM H., *These are the platformsthathavebanned Trump and hisallies*, The Washington Post, 14 gennaio 2021

DENNET D.C., *The intentionalstance*, Cambridge, 1987

Derecho al olvido: historico fallo contra Google en elpais, Cadena3, 11 agosto 2020

DESSAUER F., *Filosofia della tecnica*, a cura diBENDISCIOLIM..Brescia, 1945

DEVEREAUX A., PENG L., *Give us a little social credit: to design or to discover personal ratings in the era of Big Data*, in *Journal of Institutional Economics*, 16, 2020, p. 369-387

DI GIOACCHINO R., STOLFI F., *Data trust per un uso equo dei dati: un approccio contro lo strapotere delle Big Tech*, Agenda Digitale, 7 luglio 2021

DIGITALEUROPE, *Report – Digital Europe'sInitialFindings on the ProposedAI Act*, Bruxelles, 6 agosto 2021, p. 4 ss.

DIJKSTRA E., *The threats to computing science*, *Statement at the ACM 1984 South Central Regional Conference*, Austin, 16-18 novembre 1984

DINI A., *Debutta in Estonia il giudice-robot: le sentenze dall'intelligenza artificiale*, Corriere delle comunicazioni, 16 maggio 2019

DINUCCI D., *Fragmented future*, in *Print*, 32, 2019, p. 220-223

DOBBE R., DEAN S., GILBERT T., KOHLI N., *A Broader View on Bias in Automated Decision-Making: Reflecting on Epistemology and Dynamics*, 2018, arXiv:1807.00553

DOCKRILL P., *Controversial AI Has Been Trained to Kill Humans in a Doom Deathmatch*, Science Alert, 1 ottobre 2016.

DOCTOROW C., *Why it is not possible to regulate robots*, in *The Guardian*, 2 aprile 2014

DOEBBLER C.F.J., *Principle of non-discrimination in international law*, Washington, 2007

DOLSO G.P., *Il principio di non discriminazione nella giurisprudenza della Corte Europea dei Diritti dell'Uomo*, Napoli, 2013

DOMARADZKI S., KHVOSTOVA M., PUPOVAC D., *Karel Vasak's Generations of Rights and the Contemporary Human Rights Discourse*, in *Human Rights Review*, 20, 2019, p. 423-443.

DONATI F., *Fake news e libertà d'informazione*, in *Medialaws – Rivista di diritto dei media*, 2, 2018, p. 445 ss.

DONATI F., *Intelligenza artificiale e giustizia*, in *Rivista AIC*, 1, 2020

DORAN D., S. SCHULZ, BESOLD T.R., *What Does Explainable AI Really Mean? A New Conceptualization of Perspectives*, 2017, http://arxiv.org/abs/1710.00794

DORMEHL L., *Revisiting the rise of A.I.: How far hasartificial intelligence come since2010?,* in *Digital trends*, 2019

DÖRR D., WEAVER R.L. (a cura di), *Perspectives on privacy: increasing regulation in the USA, Canada, Australia and European countries*, Berlino-Boston, 2014

DOSHI A.R., RAGHAVAN S., WEISS R., PETITT E., *How the supply of fake news affected consumer behavior during the 2016 US election*, 2018

DOŠILOVIĆ F.K., BRČIĆ M., HLUPIĆ N., *Explainable artificial intelligence: a survey*, in *41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2018, p. 210-215

DOUVILLE T., *Parcoursup: transparence des algorithmes locaux limitée à raison pour le Conseil constitutionnel: observations sous le Conseil constitutionnel, du 3 avril 2020, n. 2020-834 QPC, UNEF - Qualification de la décision: confirmation*, in *Dalloz IP/IT* , 2020, p. 516 ss.

DREISBACH C., KOLECK T.A., BOURNE S., BAKKEN S., *A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data*, in *International Journal of Medical Informatics*, 125, 2019, p. 37-46

DUAN Y., EDWARDS J.S., DWIVEDI Y.K., *Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda*, in *International Journal of Information Management*, 48, 2019, p. 63–71

DUARTE N., LLANSO E., LOUP A., *Mixed messages? The limits of automated social media content analysis*, Report fo the Center for Democracy & Technology, 2017, p. 7 ss

DWORKIN R., *Takingrightsseriously*, Cambridge (US), 1977

DZINDOLET M., PETERSON S., POMRANKY R.A., PIERCE L.G., BECK H., *The role of trust in automation reliance*, in *International Journal of Human-Computer Studies*, 58, 6, 2003, p. 697–718

EASTERBROOK F.H., *Cyberspace and the law of the horse*, in *1996 University of Chicago Legal Forum*, 1996, p. 207-216.

EDPB, *Guidelines 05/2020 on consent under Regulation 2016/679*, 4 maggio 2020

EDWARDS L., VEALE M., *Enslaving the algorithm: from a "right to an explanation" to a "right to better decisions"?*, in *IEEE Security & Privacy*, May-June 2018, p. 46-54

EFREN RIOS VEGA L., SCAFFARDI L., SPIGNO I. (a cura di), *I diritti fondamentali nell'era della Digital Mass Surveillance*, Napoli, 2021

EL HAROUN Z., *Digital rights activists accuse Facebook of anti-palestinian bias*, Reuters, 3 novembre 2021

EL-DEMERY A.M., *The arab charter of human rights: a voice for sharia in the modern world*, Indianapolis, 2015

ELLIOT A.F., *China is banning people with bad 'social credit' from using planes and trains, in The Telegraph, 19 maggio 2018*

ELLUL J., *La technique ou l'enjeudu siècle*, Parigi, 1954

ELLUL J., *Le Systèmetechnician*, Parigi, 1977

ENDSLEY M.R., *Automation and Situation Awareness*, in PARASURAMAN R., MOULOUA M., *Automation and Human Performance: Theory and Applications*, Boca Raton (US), 1996

ENGDAHL S., *Amendment XIV: Equal Protection*, in *Constitutional amendments: beyond the bill of rights*, Farmington Hills, 2009

ENGELMANN S., CHEN M., FISCHER F., KAO C.Y., GROSSKLAGS J., *Clear Sanctions, Vague Rewards: How China's Social Credit System Currently Defines Good and Bad Behavior*, in *Proceedings of the Conference on Fairness, Accountability, and Transparency – ACM*, 2019, p. 69-78

ENGSTROM E., FEAMSTER N., *The limits of filtering: a look at the functionality and shortcomings of content detection tools*, Report - Engine, 2017, http://www.engine.is/the-limits-of-filtering/, p. 11 ss.

EPSTEIN R., ROBERT G., BEBER G. (a cura di), *Parsing the Turing Test*, Dordrecht, 2008

ERB B., *Artificial intelligence & the theory of mind*, in *Seminar Cognitition& Emotion V*, 2016, https://bit.ly/3CVhMyE (3 agosto 2022)

ERGEC R., HAPPOLD M., *Protection européenne et internationaledesdroits de l'homme*, Bruxelles, 2014

ERIKA P. ET AL., *Triage decision-making at the time of COVID-19 infection: the Piacenza strategy*, in *Internal and Emergency Medicine*, 15, 5, 2020

ESPOSITO C., *Eguaglianza e giustizia nell'art. 3 della Costituzione*, in ESPOSITO C., *La Costituzione italiana. Saggi*, Padova, 1954

ESPOSITO C., *La libertà di manifestazione del pensiero nell'ordinamento italiano*, Milano, 1958

ESTEVA A., KUPREL B., NOVOA R.A., KO J., SWETTER S.M., BLAU H.M., *Dermatologist-level classification of skin cancer with deep neural networks*, in *Nature*, 542, 7639, 2017, p. 115-118

EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment*, 2016

EUROPEAN COMMISSION HIGH LEVEL EXPERT GROUP ON AI, *A definition of Artificial Intelligence: main capabilities and scientific discipline*, 8 april 2019

EUROPEAN COMMISSION HIGH LEVEL EXPERT GROUP ON AI, *Ethics Guidelines for Trustworthy Artificial Intelligence*, 8 april 2019

EUROPEAN DATA PROTECTION BOARD, *Guidelines*2/2019 o*n the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects*, 9 aprile 2019

EUROPEAN DATA PROTECTION BOARD, *Guidelines*5/2019 *on thecriteriaoftherighttobeforgotten in thesearchenginescasesunderthe GDPR*, 7 luglio 2019

EUROPEAN INVESTMENT BANK, *Who is prepared for the new digital age? - Evidence from the EIB Investment Survey*, 2020, doi:10.2867/03951

EUROPEAN PARLIAMENT -DIRECTORATE GENERAL FOR INTERNAL POLICIES, *Study -Mapping smart cities in the EU*, gennaio 2014

EUROPEAN PARLIAMENTARY RESEARCH SERVICE, *Tackling deepfakes in European policy*, luglio 2021

EUVSDISINFO, *Report - Repeating a liedoesnot make ittrue*, 9 aprile 2020

EVANS D.S., SCHMALENSEE R.L., *Matchmakers: the new economy of multisided platforms*, Harvard, 2016

EVANS T.G., *A program for the solution of geometric analogy intelligence test questions*, in MINSKY M., *Semantic Information Processing*, Cambridge (US) 1968, p. 271-353

EVELYN E., WATSON P., *EU Anti-DiscriminationLaw*, Oxford, 2012

Exec. Order n. 13859 of Feb 11, 2019, *Maintaining American leadership in artificial intelligence*.

FABIANO N., *ePrivacy, a che punto siamo? Ecco lo stato dell'arte*, in *Agenda digitale*, 1 marzo 2022

FABIO R., *L'insostenibile complessità del processo: quale giustizia per gli small claims*, in MONATERI P.G., SOMMA A., *Patrimonio, persona e nuove tecniche di "governo del diritto". Incentivi, premi, sanzioni*, Napoli, 2009, p. 677-704

FACEBOOK OVERSIGHT BOARD, Case decision 2021-001-FB-FBR, 5 maggio 2021

FACEBOOK, *Facebook corporate human rights policy*, 16 marzo 2021

FACEBOOK, *Ourpreparationsahead of inauguration day*, 11 gennaio 2020

FACEBOOK, *Update on new Zeland*, 18 marzo 2019

FAGAN F., LEVMORE S., *The Impact of Artificial Intelligence on Rules, Standards, and JudicialDiscretion*, in *California Law Review*, 1, 2019, p. 1-37

FAILS J.A., OLSEN D.R., *Interactive machine learning*, in *IUI '03: Proceedings of the 8th international conference on Intelligent user interfaces*, 2003, p. 39-45.

FALLETTA P., *Controlli e responsabilità dei "social network" sui discorsi d'odio "online"*, in *MediaLaws*, 1, 2020, p. 146 ss.

FALLOT J., *Marx e la questione delle macchine*, Firenze, 1971.

FAN X., FENG X., DONG Y., HOU H., *COVID-19 CT image recognition algorithm based on transformer and CNN*, in *Displays*, 72, 2022

FASAN M., *Intelligenzaartificiale e pluralismo: usodelletecniche di profilazionenellospaziopubblicodemocratico*, in *BioLaw Journal – Rivista di Biodiritto*, 1, 2019, p. 107 ss.

FASAN M., *L'intelligenza artificiale nella dimensione giudiziaria. Primi profili giuridici e spunti dall'esperienza francese per una disciplina dell'AI nel settore della giustizia* in *Gruppo di Pisa*, Quaderno monografico n. 3, 2021, 325-339

FASCIGLIONE M., *Gig economy e diritti fondamentali sul lavoro in una recente sentenza del Tribunale di Bologna*, in *IRiSS*, 9 febbraio 2021, https://bit.ly/3OuXm1M (21 maggio 2022).

FEDERICO F., MARCUCCI J., BEVILACQUA M., MARCHETTI D.J., *Rapporto 2/2021 –L'impiego dell'IA nell'attività di Banca d'Italia*, in *BioLaw Journal – Rivista di BioDiritto*, 4, 2021, p. 229 ss.

FEIGENBAUM E.A., BUCHANAN B., LEDERBERG J., *On generality and problem solving: a case study using the DENDRAL program*, Stanford Artificial Intelligence Project – Computer Science Dept. Report n. CS176

FEIGENBAUM E.A., MCCORDUCK P., *The fifth generation: artificial intelligence and Japan's computer challenge to the world*, Boston, 1983

FENG Y., DUAN Q., CHEN X., YAKKALI S.S., WANG J., *Space cooling energy usage prediction based on utility data for residential buildings using machine learning methods*, in *Applied Energy*, 291, 2021, p. 116814 ss.

FERGUSON A.G., *Policing Predictive Policing*, in *Washington University Law Review*, 94, 5, 2016, p. 1109 ss.

FERNANDEZ A., *Artificial Intelligence in Financial Services*, Banco de Espana – EconomicBulletin, 2, 2019.

FERRAIOLI L., *Diritti fondamentali. Un dibattito teorico*, Bari, 2001

FERRAIOLI L., *La democrazia attraverso i diritti*, Bari, 2013

FERRARA E. ET AL., *The Rise of Social Bots*, in *Communications of the ACM*, 59, 7, p. 96–104

FERRARESE M.R., *Privatizzazioni, poteri invisibili e infrastrutture giuridiche globali*, in *Diritto pubblico*, 3, 2021, p. 871-892

FERRARI G.F. (A CURA DI), *Le smart cities al tempo dellaresilienza*, Milano, 2021

FERREIRA F.L., BOTA D.P., BROSS A., MÉLOT C., VINCENT J.L., *Serial Evaluation of the SOFA Score to Predict Outcome in Critically Ill Patients*, in *Journal of American Medical Association*, 286, 14, 2001

FERRUA P., *Il giusto processo*, Bologna, 2005

FINK C., *Dangerous speech, anti-muslim violence, and Facebook in Myanmar*, in *Journal of International Affairs*, 71, 1, 5, 2018, p. 43–52

FINOCCHIARO G., *Digital Services Act: la ridefinizione della limitata responsabilità del provider e il ruolo dell'anonimato*, in *MediaLaws*, 12 gennaio 2021

FINOCCHIARO G., *Identità personale (diritto alla)*, in *Digesto delle discipine privatistiche*, aggiornamento 2010, p. 721-738

FINOCCHIARO G., *Intelligenzaartificiale e protezionedeidatipersonali*, in *Giurisprudenzaitaliana*, 2019, p. 1670

FINOCCHIARO G., *La protezione dei dati personali e la tutela dell'identità*, in *Diritto di Internet*, Bologna, 2020, p. 151-183

FIORAVANTI M., *Costituzionalismo. La storia, le teorie, i testi*, Roma, 2018.

FISCUS R.J., *The constitutional logic of affirmative action*, Durham-Londra, 1992

FITZSIMMONS J., *Information technology and the third industrial revolution*, in *The Electronic Library*, 12, 5, p. 295-297

FJELLAND R., *Why general artificial intelligence will not be realized*, in *Humanities and social sciences communications*, 7, 10, 2020.

FLORENZANO D., *Il principio costituzionale di eguaglianza*, in CORTESE F., BORGONOVO RE D., FLORENZANO D., *Diritti inviolabili, doveri di solidarietà e principio di eguaglianza*, 2015, p. 103 ss.

FLORIDI L., *Infosfera – filosofia ed etica dell'informazione*, Torino, 2009

FLORIDI L., *The philosophy of information*, Oxford, 2011

FLORIDI L., *The ethics of information*, Oxford, 2013

FLORIDI L., *The fourth revolution – How the Infosphere is reshaping human reality*, Oxford, 2014

FLORIDI L. (a cura di),*The Onlife Manifesto*, Cham, 2015

FLORIDI L., COWLS 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*, 2018

FLORIDI L., *The logic of information*, Oxford, 2019

FLORIO F., *Didattica a distanza, il digital divide si registra in tutti i paesi*, Orizzonte scuola, 31 maggio 2020

FLORIO F., *Torna la didattica a distanza, ma il digital divide taglia fuori 300 mila studenti: «Rischio crisi educativa»,* Open, 3 novembre 2020

FLYNN S., *How Natural Language Processing (NLP) AI Is Used in Law*, Law Technology Today, 9 giugno 2019

FOÀ S., *Pubblici poteri e contrasto alle fake news. Verso l'effettività dei diritti aletici?*, in *Federalismi.it*, 11, 2020, p. 250 ss.

FODOR J., *How the mind works: what we still don't know*, in *Daedalus*, 2006, 135, 3, p. 86-94

FØLSTAD A., BRANDTZÆG P.B., *Chatbots and the new world of HC*, in *Interactions*, 24, 4, 2017, p. 41-42

FORMENTIN S., VON HEUSDEN K., KARIMI A., *Model-based and data-driven model-reference control: a comparative analysis*, in *2013 European Control Conference (ECC)*, 2013, doi:10.23919/ECC.2013.6669388

FORMICA F., *Odissea Spid: tre soluzioni per ottenerlo in modo rapido*, la Repubblica, 26 marzo 2022

FOX J., *An unlikable truth: social media like buttons are designed to be addictive. They're impacting our ability to think rationally*, in *Index of censorship*, 47, 3, 2018

FRANSSEN M.P.M., LOCKHORST G.J., VAN DE POEL I., *Philosophy of technology*, in ZALTA E.N. (a cura di), *The Stanford Encyclopedia of Philosophy*, 2018

FRASER H., COIERA E., WONG D., *Safety of patient-facingdigitalsymptomcheckers*, in *The Lancet*, 392, 10161, 2018, p. 2263-2264

FREEMAN K., *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis*, in *North Carolina Journal of Law & Technology*, 18, 5, 2016, p. 75 ss.

FREGE F.L.G., *Begriffsschrift, eine der arithmetischennachgebildeteFormelsprache des reinenDenkens*, 1879, trad. inglese a cura di BAUER-MENGELBERG S., *Concept Script, a formal language of pure thought modelled upon that of arithmetic*, in VAN HEIJENOORT J. (a cura di), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, Cambridge (US), 1967

FRESCHI C., FERRARI V., MELFI F., FERRARI M., MOSCA F., CUSCHIERI A., *Technical review of the da Vinci surgical telemanipulator*, in *The International Journal of Medical Robotics and Computer Assisted Surgery*, 9, 4, p. 396-406

FRIED L.P., PACCAUD P., *Editorial: The Public Health Needs for an Ageing Society*, in *Public Health Reviews*, 32, 2, 2010, p. 351-355

FRIEDENBERG J., SILVERMAN G., *Cognitive science: an introduction to the science of the mind (3 ed.)*, Los Angeles, 2016

FRIEDMAN B., NISSENBAUM H., *Bias in computer systems*, in *ACM Transactions on Information Systems*, 3, 1996, doi.org/10.1145/230538.230561

FRONZA E., *"Code is Law". Note a margine del volume di Antoine Garapon e Jean Lassègue, Justice Digitale. Révolutiongraphique et ruptureanthropologique, Puf, Paris*, in *Diritto Penale Contemporaneo*, 11 dicembre 2018, https://bit.ly/3HsXBbA (14 maggio 2022)

FROSINI T.E., *Internet come ordinamento giuridico*, in *Percorsi costituzionali*, 1, 2014, p. 262 ss.

FROSINI T.E., *No news is fake news*, in *DPCE*, 4, 2017, p. 4 ss.

FROSINI T.E., *Liberté Egalité Internet*, Napoli, 2019

FROSINI T.E., *La privacy nell'era dell'intelligenza artificiale*, in *DPCE Online*, 51, 1, 2022

FROSIO G. F., *Reforming intermediary liability in the platform economy: a European digital single market strategy*, in *Northwestern University Law Review*, 112, 2017, p. 19 ss

FROSIO G. F., *Why keep a dog and bark yourself? From intermediary liability to responsibility*, in *International Journal of Law and Information Technology*, 26, 2018, p. 1-33

FULLER S.H., MILLET L.I., *The future of computing performance. Game over or nextlevel?*, Washington, 2011

G20 MINISTERIAL STATEMENT ON TRADE AND DIGITAL ECONOMY, 9 giugno 2019

GABBAT A., *IBM computer Watson wins Jeopardy clash*, The Guardian, 17 febbraio 2011

GABBATT A., *Claim of anti-conservative bias by social media firms is baseless, report finds*, The Guardian, 1 febbraio 2021

GACUTAN J., SELVADURAI N., *A statutory right to explanation for decisions generated using artificial intelligence*, in *International Journal of Law and Information Technology*, 28, 2020, p. 193-216

GADAMER H.G., *Verità e metodo*, Tubinga, 1960

GAJDA A., *Privacy, press and the right to be forgotten in the United States*, in *Washington Law Review*, 93, 201, 2018

GALETTA D.U., CORVALÁN J.C., *Intelligenza artificiale per una pubblica amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *Federalismi.it*, 3, 2, 2019

GALLAGHER D.A., *Free speech on the line: modern technology and the First Amendment*, in *Comm Law Conspectus: Journal of Communications Law & Policy*, 3, 2, 1995, p. 197-206

GALLI G., *Difesa dell'imputato e speditezza del processo: dalla Costituzione alle leggi dell'emergenza*, Milano, 1982

GALLONE G., OROFINO A.G., *L'intelligenza artificiale al servizio delle funzioni amministrative: profili problematici e spunti di riflessione. Nota a sent. Cons. Stato sez. VI 4 febbraio 2020 n. 881*,in*Giurisprudenza italiana*, 2020, 7, p. 1738-1748

GAMBARO A., *Falsa luce agli occhi del pubblico (false light in the public eye)*, in *Riv. dir. civ.*, 1981, p. 84-135

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, Decisione 10 novembre 2004, https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/1116068

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Parere alla Provincia autonoma di Trento su uno schema di regolamento concernente la medicina di iniziativa nel servizio sanitario provinciale*, 1 ottobre 2020

GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, *Linee guida cookies e altri strumenti di tracciamento*, 10 giugno 2021

GARAPON A., LASSEGUE J., *Justice Digitale: révolutiongraphique et ruptureanthropologique*, Parigi, 2018

GARAPON A.,LASSÈGUE J., *Justice digitale,* Parigi, 2018

GARBADE M.J., *A simple introduction to Natural Language Processing*, in *Becoming humans: Artificial Intelligence magazine*, 15 ottobre 2018

GARCÍA-MAGARIÑO I., MUTTUKRISHNAN R., LLORET J., *Human-Centric AI for Trustworthy IoT Systems With Explainable Multilayer Perceptrons*, in *Access*, 7, 2019, p. 125562–125574

GARGEYA R.,LENG T., *Automated Identification of Diabetic Retinopathy Using Deep Learning*, in *Ophthalmology*, 124, 7, 2017, p. 962-969

GARTENBERG C., *Big tech's 2021 earningswere off the chart*, The Verge, 11 febbraio 2022

GASKIN J., JENKINS J., MESERVY T., STEFFEN J., PAYNE K., *Using Wearable Devices for Non-invasive, Inexpensive Physiological Data Collection*, in *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017, http://hdl.handle.net/10125/41221

GASPERETTI M., *Coronavirus, il piano anti-pandemia che l'Italia non ha seguito*, Corriere della Sera, 28 marzo 2020

Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report, 2021, https://stanford.io/3KO5R7K

GAWER A., *Platforms, markets and innovation*, Cheltenham, 2010

GELLER T., *Overcoming the Uncanny Valley*, in *IEEE Computer Graphics and Applications*, 28, 4, p. 11 ss.

GELLERT R., *The risk-based approach to data protection*, New York, 2020

GELLERT R., *The role of the risk-based approach in the General Data Protection Regulation and in the European Commission's proposed Artificial Intelligence Act: business as usual?*, in *Journal of Ethics and Legal Technologies*, 3, 2, 2021, p. 15-33

GELLERT R., *Understanding the notion of risk in the General Data Protection Regulation*, in *Computer Law & Security Review*, 34, 2, 2018, p. 279-288.

GENOVESE U., ZOJA R., FERRARIO A., SERPETTI A., MARIOTTI P., *La medicina difensiva. Questioni giuridiche, assicurative, medico-legali*, Bologna, 2011

GENTILI A., *Il principio di non discriminazione nei rapporti civili*, in *Rivista critica del diritto privato*, 2, 2009, p. 207-231

GESLEVICHPACKIN N., LEV-ARETZ Y., *Learning algorithms and discrimination*, in BARFIELD W., PAGALLO U. (A CURA DI),*Research Handbook on the Law of Artificial Intelligence*, Cheltenham-Northampton (MA), 2018, p. 109 ss.

GHODSELAHI A., AMIRMADHI A., *Application of artificial intelligence techniques for credit risk evaluation*, in *International journal of modeling and optimization*, 1, 3, 2011, p. 243 ss.

GIALUZ M., *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Diritto Penale Contemporaneo*, 2019

GIL M., ALBERT M., FONS J., PELECHANO V., *Engineering human-in-the-loop interactions in cyber-physical systems*, in *Information and Software Technology*, 126, 2020

GILLESPIE T., *Platforms are not intermediaries*, in *Georgetown Law Technology Review*, 2, 2, p. 198-216

GILLESPIE T.,*Custodians of the Internet: Platforms, Content Moderation, and the HiddenDecisionsthatShape Social Media*, New Haven, 2018

GINZBURG R.B., *Interpretations of the equalprotectionclause*, in *Harvard Journal of Law& Public Policy*, 9, 1, 1986, p. 41-46

GIORGI M.C., *L'uguaglianza sostanziale nel lungo dibattito costituente*, in *Rivista Trimestrale di Diritto Pubblico*, 1, 2018, p. 9-43

GIORGIS A., *Art. 3 c. 2*, in BIFULCO R., CELOTTO A., *Commentario alla Costituzione*, I, Torino, 2006, p. 88-113

GIORGIS A., *La costituzionalizzazione dei diritti all'eguaglianza sostanziale*, Napoli, 1999

GLASER A., *When Robots Make Us Angry, Humans Pay the Price*, Slate, 14 settembre 2017

GLAUNER P., *An Assessment of the AI Regulation Proposed by the European Commission*, 26 maggio 2021, http://arxiv.org/abs/2105.15133

GLIKSON E., WILLIAMS WOOLLEY A., *Human Trust in Artificial Intelligence: Review of Empirical Research*, in *Academy of Management Annals* 14, 2, 2020, p. 627–660

GLINSKY V.A., GLINSKII A.B., STEPHENSON A.J., HOFFMAN R.H., GERALD W.L., *Gene expression profiling predicts clinical outcome of prostate cancer*, in *Journal of Clinical Investigation*, 113, 6, 2004, p. 913-923

GLOBOCNIK J., *The Right to Be Forgotten is Taking Shape: CJEU Judgments in GC and Others (C-136/17) and Google v CNIL (C-507/17)*, in *GRUR International – Journal of European and International IP Law*, 69, 4, p. 380-388

GOEL V., *Facebook scrambles to police content amid rapid growth*, The New York Times, 3 maggio 2017

GOLDING-YOUNG S., *Facebook's discrimination against the LGBT community*, ACLU, 24 settembre 2020

GOLIA JR. A., *L'antifascismo della Costituzione italiana alla prova degli spazi giuridici digitali. Considerazioni su partecipazione politica, libertà di espressione "online" e democrazia (non) protetta in "CasaPound c. Facebook" e "Forza Nuova c. Facebook"*, in *Federalismi.it*, 18, 2020, p. 134 ss.

GONCALVES JR. A.F., *Etica e sociedadetecnológicasegundo a filosofia de Ortega y Gasset*, in *Reflexao*, Campinas, 31, 89, p. 25-39

GONCALVES M.E., *The risk-based approach under the new EU data protection regulation: a critical perspective*, in *Journal of risk research*, 23, 2, 2020, p. 139-152

GONZÁLEZ QUIRÓS J.L., *La meditación de Ortega sobre la técnica y las tecnologías digitales*, in *Revista de estudios orteguianos*, 2006, 12-13, p. 95 ss.

GOODMAN B., FLAXMAN S., *European Union regulations on algorithmic decision-making and a "right to explanation"*, in *AI Magazine*, 2017, p. 50-57

GOOGLE, *AI ExplainabilityWhitepaper*, 2020

GORWA R., BINNS R., KATZENBACH C., *Algorithmic content moderation: Technical and political challenges in the automation of platform governance*, in *Big Data & Society*, 7, 1, 2020

GOSS R., *Criminal fair trailrights: article 6 of the European Convention of Human Rights*, Londra, 2016

GOURDEAU L. ET AL., *Deep learning of chest X-rays can predict mechanical ventilation outcome in ICU-admitted COVID-19 patients*, in *Scientific Reports*, 12, 1, 2022

GRAHAM D., *An internet in your head*, NewYork, 2021

GRAHAM H., *The Civil Rights Era: Origins and Development of National Policy 1960–1972*, New York, 1990

GRANDINETTI O., *Facebook vs. CasaPound e Forza Nuova, ovvero la disattivazione di pagine social e le insidie della disciplina multilivello dei diritti fondamentali*, in *MediaLaws*, 1, 2021, p. 173 ss

GREEN B.P., *Artificial intelligence, decision-making and moral deskilling*, in *Markkulla center for applied ethics*, 15 maggio 2019, https://bit.ly/3SWwjiB (12 settembre 2022)

GREEN T., *How to stop fearing black box AI and love the robot-ruled future*, TNW, 8 dicembre2017, https://bit.ly/3BiyarT (2 agosto 2022)

GREGORY R.F., *The Civil Rights Act and the Battle to End Workplace Discrimination*, Lanham, 2014

GRIECO L., *Informazioni e accesso ai dati personali – artt. 13, 14, 15*, in BOLOGNINI L., PELINO E. (A CURA DI), *Il codice dela disciplina privacy*, Milano, 2019, p. 147 ss.

GRINBERG N. ET AL., *Fake News on Twitter during the 2016 U.S. Presidential Election*, in *Science*, 363, 6425, 2019, p. 374–378

GROSSI P.F., *Introduzione allo studio dei diritti inviolabili nella Costituzione italiana*, Padova, 1972

GUASTINI R., *Interpretare e argomentare*, in CICU A., MESSINEO F., MENGONI L., *Trattato di diritto commerciale*, Milano, XIV, 2021

GUI L., *L'utente che non c'è. Emarginazione grave, persone senza dimora e servizi sociali*, Milano, 1995

GUIDOTTI, R. MONREALE A., PEDRESCHI D., *The AI black-box explanation problem*, in *ERCIM News*, 116, 2019, p. 12-13

GUIDOTTI R., MONREALE A., RUGGIERI S., TURINI F., GIANOTTI F., PEDRESCHI D., *A survey of methods for explaining black box models*, in *ACM Computing Surveys*, 51, 5, 2018, p. 93 ss.

GUNNING D., STEFIK M., CHOI J., MILLER T., STUMPF S., YANG G., *XAI - Explainable artificial intelligence*, in *Science Robotics*, 4, 37, 2019, https://doi.org/10.1126/scirobotics.aay7120

GUO L.N., LEE M.S., KASSAMALI B., MITA C., NAMBUDIRI V.E., *Bias in, bias out: underreporting and underrepresentation of diverse skintypes in machine learning research for*

skincancerdetection—A scoping review, in *Journal of the American Academy of Dermatology*, 87, 1, 2022

GURUMOORTHY S., RAO B.N., GAO X.Z., *Cognitive sciences and artificial intelligence*, Singapore, 2018

GUTHRIE C.P., RACHLINSKI J.J., WISTRICH A.J., *Inside the Judicial Mind*, in *Cornell Law Faculty Publications,* 814*,* 2001, http://www.ssrn.com/abstract=257634

HAAG A.M., BOYES A., CHENG J., MACNEIL A., WIROVE R., *An introduction to the issues of cross-cultural assessment inspired by Ewert v. Canada*, in *Journal of Threat Assessment and Management, 3, 2, 2016,* p. 65–75

HABAYEB A., *Explainable AI isn'tenough; we need under standable AI*, in *Techopedia.com*, https://bit.ly/3CWvKA8

HACKER P., *A legal framework for AI training data – from first principles to the Artificial Intelligence Act*, 2021

HADFIELD-MENELL D., DRAGAN A., ABBEEL P., RUSSELL S., *The Off-Switch Game*, 2016, https://arxiv.org/abs/1611.08219v3

HAN E.R. ET AL., *Medical education trends for future physicians in the era of advanced technology and artificial intelligence: an integrative review*, in *BMC MedicalEducation*, 19, 1, 2019

HAN-WEI L., CHING-FU L., YU-JIE C., *Beyond State v Loomis: Artificial Intelligence, Government Algorithmization and Accountability*, in *International Journal of Law and Information Technology*, 27, 2, p. 122–141

HAO K., *Deepfake porn is ruining women's lives. Now the law is finally banning it*, in *MIT Technology Review*, 12 febbraio 2021

HAO K., *Doctors are using AI to triage covid-19 patients. The tools may be here to stay*, in *MIT Technology Review*, 23 aprile 2020

HAO K., *Is China's social credit system as Orwellian as it sounds?*, in *MIT Technology Review*, 26 febbraio 2018

HAO K., *Porn sites won't take down nonconsensual deepfakes*, Wired, 30 agosto 2020

HARARI Y.N.,*Homo deus: a brief history of tomorrow*, Gerusalemme, 2015

HARNAD S.R., *Minds, machines and Searle*, in *Journal of theoretical and experimental artificial intelligence*, 1, 1989, p. 5-25

HARNAD S.R., *Other bodies, other minds: a machine incarnation of an old philosophical problem*, in *Minds and Machines*, 1, 1991, p. 43-54

HARNAD S.R., *The Turing test is not a trick: Turing indistinguishability is a scientific criterion*, in *ACM SIGART Bulletin*, 3, 4, 1992, p. 9-10

HARRIS S.J., ARAMBULA-COSIO F., MEI Q., HIBBERD R.D., DAVIES B.L., WICKHAM J.E.A., *The Probot—an active robot for prostate resection*, in *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 211, 4, 1997, p. 317-325

Hashing, Techopedia, 2021, https://bit.ly/3mc9mJv

HAUGELAND J., *Artificial intelligence: the very idea*, Cambridge (US)-Londra, 1985

HAUSMAN D.M., WELCH B., *Debate: to nudge or not to nudge*, in *Journal of PoliticalPhilosophy*, 18, 1, p. 123-136

HAYAT A., *Law Firms and Their Tech: Clifford Chance, Ashurst, Freshfields, DLA Piper, Mayer Brown*, New Law Academy, agosto 2021, https://bit.ly/3QOznM

HAYAT A., *The future of the law: White & Case on artificial intelligence*, Chambers Associate, novembre 2019, https://bit.ly/3BSHgvI

HAYES-ROTH F., *Rule-based systems*, in *Communications of the ACM*, 28, 9, 1985 p. 921–932, doi:10.1145/4284.4286

MARTIN HEIDEGGER, *Essere e tempo*, Halle, 1927

HEIDEGGER M., a cura di F. SOLLAZZO, *La questione della tecnica*, Firenze, 2017

HEIKKILÄ M., *A Quick Guide to the MostImportant AI LawYou'veNever Heard of*, in *MIT Technology Review*, 2022

HEKLER A. ET AL., *Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images*, in *European Journal of Cancer*, 118, 2019, p. 91 ss.

HELLER J., *Syphilisvictims in U.S. study wentuntreated for 40 years*, New York Times, 26 luglio 1972

HEMMENDINGER D., *LISP – computer language*, in *Encyclopedia Britannica*, 2016, https://www.britannica.com/technology/LISP-computer-language

HENRY B., *Dal golem ai cyborg. Trasmigrazioni nell'immaginario*, Livorno, 2013.

HERN A., *Revealed: catastrophic effects of working as a Facebook moderator*, The Guardian, 17 settembre 2019

HERN K. RAWLINSON A., *Facebook bans Britain First and its leaders*, in *The Guardian (online)*, 14 marzo 2018

HERREROS B., GELLA P., DE ASUA D.R., *Triage during the Covid-19 epidemic in Spain: better and worseethicalarguments*, in *Journal of Medical Ethics*, 46, 7, 2020

HEUNG A., WEBER R. H., *Internet governance and the responsibility of Internet Service Providers*, in *Wisconsin International Law Journal*, 26, 2, 2008, p. 403-477

HIGH LEVEL EXPERT GROUP ON AI della Commissione UE, *Policy and Investment Recommendations for Trustworthy AI*, 26 giugno 2019

HILL E., WOLFE J., *Ewert v. Canada: Shining Light on Corrections and Indigenous People*, in *The Supreme Court Law Review: Osgoode's Annual Constitutional Cases Conference* , 94, 15, 2020, p. 391-413

HILL J., RANDOLPH FORD W., FARRERAS I.G., *Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations*, in *Computers in Human Behavior*, 2, 2015, p. 245-250

HINES M., *I Smell a Bot: California's S.B. 1001, Free Speech, and the Future of Bot Regulation*, in *Houston Law Review*, 57, 2, 2019, p. 40 ss.

HJORTH L., HINTON S., *Understanding social media (2nd ed.)*, Londra, 2019.

HOCHREITER S., SCHMIDHUBER J., *Long short-term memory*, in *Neural Computation*, 9, 8, 1997, p. 1735-1780

HOFFMAN S., *Managing the state: social credit, surveillance and CCP's plan for China*, in WRIGHT N. (a cura di), *AI, China, Russia and the global order: technological, political, global and creative*, Maxwell AFB, 2019, p. 48-55

HOOKER S., *Moving beyond "algorithmic bias is a data problem"*, in *Patterns*, 2, 4, 2021, https://doi.org/10.1016/j.patter.2021.100241

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, p. 2554–2558

HORNING R.A., *The first amendment right to a public forum*, in *Duke Law Journal*, 1969, 931 ss.; R.C. POST, *Between governance and management: the history and theory of the public forum*, in *UCLA Law Review*, 34, 1987, p. 1713 ss.

HORTON R., *African Traditional Thought and Western Science*, in *Africa*, 37, 1, 1967, p. 50-71

HOUSER K., BAGBY J.W., *The data trust solution to data sharing problems*, in *Vanderbilt Journal of Entertainment & Technology Law*, 2022, http://dx.doi.org/10.2139/ssrn.4050593

HRYNIEWSKAW,,BOMBIŃSKI P., SZATKOWSKI P. ET AL., *Do not repeat these mistakes -- a critical appraisal of applications of explainable artificial intelligence for image based COVID-19 detection*, 2020, https://europepmc.org/article/PPR/PPR274054 (arXIv preprint)

HU Z., TANG J., WANG Z., ZHANG K., ZHANG L., SUN Q., *Deep learning for image-based cancer detection and diagnosis − A survey*, in *Pattern recognition*, 83, 2018, p. 134-149

HUDSON D.L., *Public forum doctrine*, in *The First Amendment Encyclopedia*, 2020

HUDSON D.L., *The Fourteenth Amendment. Equal protection under the law*, Berkeley, 2002

HUDSON P., *The industrial revolution*, Londra, 2014, p. 166 ss.

HUNT M.R., SINDING C., SCHWARTZ L., *Tragic choices in humanitarian health work*, in *The Journal of clinical ethics*, 23, 4, 2012

HUNTINGTON S.P., *The thirdwawe: democratization in the late twentiethcentury*, Norman, 1991

HUTCHINS W.J., *Machine translation: a brief history*, in KOERNER E.F.K., ASHER R.E. (a cura di),*Concise history of the language sciences. From the Sumerians to the cognitivists*, Amsterdam, 1995, p. 431-445

HUTTER EPSTEIN R.*, Can a smartwatch saveyourlife?*, The New York Times, 26 luglio 2021

HWANG T., *Computational power and the social impact of artificial intelligence*, 2018, http://dx.doi.org/10.2139/ssrn.3147971 (1febbraio 2021)

I.R. PAVONE, *La convenzione europea sulla biomedicina*, Milano, 2009.

IMARISIO M., *Coronavirus, il medico di Bergamo: «Negli ospedali siamo come in guerra. A tutti dico: state a casa»*, Corriere della Sera, 9 marzo 2020

Indiansfight for the "right to be forgotten online", AlJazeera, 16 marzo 2022 https://bit.ly/3iVLHLP (20 marzo 2022)

ISA F.G., DE FEYTER K., *International protection of human rights: achievements and challenges*, Bilbao, 2006

ISAAC W., *Hope, Hype e Fear, The Promise and Potential Pitfalls of Artificial Intelligence in Criminal Justice,* in *Ohio State Journal of Criminal Law,* 15, 2017

ISHII K., *Comparative legal study on privacy and personal data protection for robots equipped with artificial intelligence: looking at functional and technological aspects*, in *AI & Society*, 34, 2019, p. 509-533, DOI 10.1007/s00146-017-0758-8

ISLAM M.R., AHMED M.U., BARUA S., BEGUM S., *A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks*, in *Applied Sciences*, 12, 3, 2022, p. 1353 ss.

ITALIA V., *Il bilanciamento nelle leggi*, Milano, 2016

J. HUTCHINS, *The first public demonstration of machine translation: the Georgetown-IBM system,* 7 gennaio 1954, https://bit.ly/39eWnR4

J. SLANEY, *Blocks World revisited*, in *Artificial Intelligence*, 125, 1-2, 2001, p. 119-153

JAGGI S., *State action doctrine*, in *Max Planck Encyclopedia of Comparative Constitutional Law*, 2017, https://oxcon.ouplaw.com/view/10.1093/law-mpeccol/law-mpeccol-e473

JAIN G., MITTAL D., THAKUR D., MITTAL M.K., *A deep learning approach to detect Covid-19 coronavirus with X-Ray images*, in*Biocybernetics and Biomedical Engineering*, 40, 4, 2020

Japan pushingahead with Society 5.0 to overcomechronic social challenges, *UNESCO Science Report*, 21 febbraio 2019

JARRAHI M.H., *Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making*, in *Business Horizons*, 61, 4, 2018, p. 577–586

JELLINEK G., *La dichiarazione dei diritti dell'uomo e del cittadino* (1985), Bari, 2002

JOBIN A., IENCA M., VAYENA E., *The global landscape of AI ethics guidelines*, in *Nature Machine Intelligence*, 1, 2019, p. 389-399

JONES M.L., *The right to a human in the loop: Political constructions of computer automation and personhood*, in *Social Studies of Science*, 47, 2, 2017, p. 216-239.

JONSSON CORNELL A., *The right to privacy*, in *The Max Planck Encyclopedia of Comparative ConstitutionalLaw*, Oxford, 2017

JOY B., *Why the future doesn't need us*, in *Wired*, 2000

JULIA L., *L'intelligence artificial n'existe pas*, Parigi, 2019

JUNG C. ET AL., *Disease-Course Adapting Machine Learning Prognostication Models in Elderly Patients Critically Ill With COVID-19: Multicenter Cohort Study With External Validation*, in *JMIR Medical Informatics*, 10, 3, 2022

JUSO S., *Motivi e motivazione nel provvedimento amministrativo*, Milano, 1960

KAHNEMAN D., SIBONY O., SUNSTEIN C.R., *Noise: a flaw in human judgment*, Boston, 2021

KAHNEMAN D., TWERSKY A., *Prospect theory: an analysis of decision under risk*, in *Econometrica*, 47, 1979, p. 263-291

KAHNEMAN D., TWERSKY A., SLOVIC P., *Judgment under uncertainity. Heuristics and biases*, Cambridge, 1982

KAHNG M., ANDREWS P.Y., KALRO A., CHAU D.H., *ActiVis: Visual Exploration of Industry-Scale Deep Neural Network Models*, in *IEEE Transactions on Visualization and Computer Graphics*, 24, 1, 2018, p. 88-97

KAMINSKI M.E., *The right to explanation, explained*, in *Berkeley Technology Law Journal,* 34, 1, 2019, p. 189 ss.

KANDEL E.R., SCHWARTZ J.H., JESSELL T.M., SIEGELBAUM S.A., HUDSPETH A.J., *Principles of neural science (Vth ed.)*, New York, 2012

KARAMANGLA S., *California hospitals face a 'war zone' of flupatients — and are setting up tents to treatthem*, Los Angeles Times, 16 gennaio 2018

KARDON B., *Is every company really an AI company?* in *Ad Age*, 2019, http://bit.ly/3gNevm6

KATZ R.V., KEGELES S.S., KRESSIN N.R., LEE GREEN B., WANG M.Q., JAMES S.A., RUSSELL S.L., CLAUDIO C., *The Tuskegee Legacy Project: willingness of minorities to participate in biomedicalresearch*, in *Journal of healthcare for the poor and underserved*, 17, 4, 2006, p. 698-715

KAUFMANN-KOHLER G., SCHULTZ T., *Online Dispute Resolution: Challenges for Contemporary Justice*, Alphen, 2004

KAUL V., ENSLIN S., GROSS S.A., *History of artificial intelligence in medicine*, in *Gastrointestinalendoscopy*, 10, 92, 4, 2020, p. 807-808

KAUL V., ESLIN S., GROSS S.A., *History of artificial intelligence in medicine*, in *Gastrointestinal Endoscopy*, 92, 4, 2020, p. 807-812

KAWA C., IANIRODAHM P.M., NIJHUIS J.F.H., GIJSELAERS W.H., *Cafeteria online: nudges for healthier food choices in a university cafeteria – a randomized online experiment*

KEDING C., MEISSNER P., *Managerial Overreliance on AI-Augmented Decision-Making Processes: How the Use of AI-Based Advisory Systems Shapes Choice Behavior in R&D Investment Decisions*, in *Technological Forecasting and Social Change*, 171, 2021

KEETON W.L., PROSSER W., *On the law of torts*, St. Paul, 1984, p. 849 ss.

KELLER D., *Making Google the censor*, in *The New York Times*, 12 giugno 2017

KELLUOGH J., *Understanding affirmative action*, Washington, 2006

KENNEDY L.W., CAPLAN J.M., PIZA E.L., *Risk Clusters, Hotspots and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies*, in *Journal of Quantitative Criminology*, 1, 2010, p. 339 ss.

KENNEY M., ZYSMAN J., *The rise of platform economy*, in *Issues in science and technology*, 2016, p. 61-69

KHOURY M.J., IOANNIDIS J.P.A., *Big data meets public health*, in *Science*, 346, 6213, 2014, p. 1054 ss.

KICKBUSCH I., PISELLI D., AGRAWAL A., BALICER R., BANNER O., ADELHARDT M. ET AL., *The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world*, in *Lancet*, 398, 10312, 2021, p. 1727-1776

KIM S.Y. ET AL., *A deep learning model for real-time mortality prediction in critically ill children*, in *Critical Care*, 23, 1, 2019

KLEINBERG J., LUDWIG J., MULLAINATHAN S., SUNSTEIN C.R., *Discrimination in the age of algorithms*, in *Journal of legal analysis*, 10, 2018, p. 113-174

KLINE R.R., *Cybernetics, Automata Studies, and the Dartmouth Conference on Artificial Intelligence*, in *IEEE Annals of the History of Computing*, 33, 4, 2011, p. 5-16

KLONICK K., *The new governors: the people, rules, and process governing online speech*, in *Harvard Law Review*, 131, 6, 2018, p. 1639 ss.

KNIGHT W., *AI's Language Problem*, in *MIT Tech. Review*, 2016

KNOX J., WILLIAMSON B., BAYNE S., *Machine behaviourism: future visions of "learnification" and "datafication" across humans and digital technologies*, in *Learning, Media & Technology*, 45, 1, 2020, p. 31-45

KÖCHLING A., WEHNER M.C., *Discriminated by an algorithm: asystematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development*, in *Business Research*, 13, 3, p. 795–848

KOHLER J., *Recht und Persönlichkeit in derKulturderGegenwart*, Stoccarda, 1914

KOSSEFF J., *The twenty-six words that created the internet*, Ithaca, 2019

KOSTA E., *Consent in European data protectionlaw*, Leiden, 2013

KOZYRKOV C., *Explainable Ai won't deliver. Here's why*, in *Medium*, 16 novembre 2020

KRAJNA A., KOVAC M., BRCIC M., SARCEVIC A., *Explainable artificial intelligence: an updated perspective*, in *45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*, 2022

KRIPPENDORFF K., *Combinatorial Explosion,* in *Web Dictionary of Cybernetics and Systems*, http://pespmc1.vub.ac.be/ASC/COMBIN_EXPLO.html

KRISHNA S., HAN T., GU A., POMBRA J., JABBARI S., WU S., *The Disagreement Problem in Explainable Machine Learning: A Practitioner's Perspective*, 2022, https://arxiv.org/abs/2202.01602

KRISHNAN S., WU E., *PALM: Machine Learning Explanations For Iterative Debugging*, in *Proceedings of the 2nd workshop on Human-In-the-Loop data analytics,* in *ACM, Chicago IL USA*, 2017, p. 1-6

KRISLOV S., *American Welfare Policy and the Supreme Court*, in *Current History*, 65, 383, 1973, p. 33–42

KRIZHEVSKY A., SUTSKEVER I., HINTON G.E., *ImageNet classification with deep convolutional neural networks*, in *NIPS*, 1, 2012

KROGH A., *What are artificial neural networks?*, in *Nature Biotechnology*, 26, 2008, p. 195-220

KRONCKE C., *Nudging towards a stable retirement*, in *Politics and the Life Sciences*, 37, 1, 2018, p. 126-129.

KRONER D.G., *The Ewert v. Canada judgment: Moving forward*, in *Journal of Threat Assessment and Management,* 3, 2, 2016, p. 122–127

KÜBLER R., PAUWELS K., MANKE K., *How Social Media Drove the 2016 US Presidential Election: A Longitudinal Topic and Platform Analysis*, Rochester Social Science Research Network, luglio 2020, https://papers.ssrn.com/abstract=3661846 (2 maggio 2022)

KUCEWICZ-CZECH E., DAMPS M., *Triage during the Covid-19 pandemic*, in *Anaesthesiology Intensive Therapy*, 52, 4, 2020, p. 312-315

KUCZERAWY A., *Intermediary liability &freedom of expression: recentdevelopments in the EU notice& action*, in *Computer Law& Security Review*, 31, 1, 2015, p. 46 ss.

KULK S., ZUIDERVEENBORGESIUS F., *Case Notes: Google Spain vs. González: did the Court forget about freedom of expression?*, in *European Journal of Risk Regulation*, 3, 2014, p. 389-398

KURNIAWAN I., POWER J.M., CANECA V.I., SOFIANTI T.D., *System Modelling and Simulation for Study of Human-Machine Collaboration Technologies Implementation on Assembly Line*, in *Proceedings of the International Conference on Engineering and Information Technology for Sustainable Industry*, Association for Computing Machinery, New York, 2020, https://doi.org/10.1145/3429789.3429799

KURZWEIL R., *The age of intelligent machines*, Cambridge, 1990

KURZWEIL R., *The age of spiritual machines*, Cambridge, 1998

KURZWEIL R., *The singularity is near*, New York, 2005

KWON J., JEON J., KIM H.M. ET AL., *Deep-learning-based out-of-hospital cardiac arrest prognostic system to predict clinical outcomes*, in *Resuscitation*, 139, 2019, p. 84-91

LA SPINA A., CAVATORTO S., *Le Autorità Indipendenti*, Bologna, 2008

LA VATTIATA F.C., *Brevi note "a caldo" sulla recente Proposta di Regolamento UE in tema di intelligenza artificiale*, in *Diritto Penale e Uomo*, 6, 2021

LAMBRECHT A., TUCKER C., *Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads*, in *Management Science*, 65, 7, 2019, p. 2966–2981

LANA A., *Anziani e tecnologia: dallo Spid ai servizi bancari un divario sempre più incolmabil*e, Corriere della Sera, 1 aprile 2022

LANCHESTER F., *Le costituzioni tedesche da Francoforte a Bonn. Introduzione e testi*, Milano, 2009

LANE T., *A short history of robotic surgery*, in *Annals of the Royal College of Surgeons of England*, 100, 6, 2018, p. 5 ss.

LANGER M., OSTER D., SPEITH T., HERMANNS H., KÄSTNER L., SCHMIDT E. ET AL., *What do we want from Explainable Artificial Intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research*, in *Artificial Intelligence*, 7, 2021

LANZA V., *Svezia, sterilizzate a forza*, la Repubblica, 25 agosto 1997

LARSON S., *Tech giants bolster collaborative fight against terrorism*, CNN Business, 26 giugno 2017

LASH K.T., *The Fourteenth Amendment and the privileges and immunities of American citizenship*, Cambridge, 2014

LASSAU N. ET AL., *Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID-19 patients*, in *Nature Communications*, 12, 1, 2021

LAUDER K., *The GermanProposal of an "Anti-Discrimination"-Law: Anticostitutional and Anti-Common Sense. A Response to Nicola Venneman*, in*GermanLaw Journal*, 3, 5, 2002

LAURO A., *Siamo tutti giornalisti? Appuntisullalibertà di informazionenell'era social*, in *MediaLaws – Rivista di dirittodei media*, 2, 2021, p. 1–24

LAVAGNA C., *Basi per uno studio delle figure giuridiche soggettive contenute nella Costituzione italiana*, Padova, 1953

LAVECCHIA A., *Deep learning in drug discovery: opportunities, challenges and future prospects*, in *Drug Discovery Today*, 24, 10, 2019, p. 2017-2032

LAVENUE L.M., MYLES J.M., SCHNEIDER A.N., *EvaluatingChina's New 'Internet Information Service AlgorithmicRecommendation Management' Regulations*, in *Finnegan*, 21 aprile 2022

LAVIOLA F., *Algoritmico, troppo algoritmico: decisioni amministrative automatizzate, protezione dei dati personali e tutela delle libertà dei cittadini alla luce della più recente giurisprudenza amministrativa*, in *Biolaw Journal*, 3, 2020, p. 389 ss.

LAVORGNA A., SUFFIA G., *La nuova proposta europea per regolamentare i Sistemi di Intelligenza Artificiale e la sua rilevanza nell'ambito della giustizia penale: un passo necessario, ma non sufficiente, nella giusta direzione*, in *Diritto Penale Contemporaneo*, 2, 2021, p. 88 ss.

LAWRENCE D.M., *Private exercise of governmentalpower*, in *Indiana Law Journal*, 61, 1986, p. 647-696

LECUN Y., *Une procedure d'apprentissage pour réseau a seuilasymmetrique (a Learning Scheme for Asymmetric Threshold Networks)*, in *Proceedings of Cognitiva 85*, Paris, 1985, p. 599-604

LEE D., *Germany'sNetzDG and the Threat to Online Free Speech*, in *Yale Law School – MFIA*, 10 ottobre 2017

LEE J., LIU C., KIM J., CHEN Z., SUN Y., ROGERS J.R. ET AL., *Deep learning for rare disease: a scoping review*, medRxiv, 2022, https://doi.org/10.1101/2022.06.29.22277046

LEE J., *Postcards from Planet Google*, The New York Times, 28 novembre 2002

LEE MYERS S., *China spins tale that the U.S. armystarted the coronavirus epidemic*, The New York Times, 13 marzo 2020

LEMMON E.J., *Beginninglogic*, Londra-Edimburgo, 1965

LEPAGE A., MARINO L., *Droits de la personnalité*, in *Recueil Dalloz*, 39, 2007, p. 2771 ss.

LESSIG L., *Code and other laws of cyberspace*, New York, 1999

LEVY J., JOTKOWITZ A., CHOWERS I., *Deskilling in Ophthalmology Is the Inevitable Controllable?*, in *Eye*, 33, 3, 2019

LI T., CUI Y., BELONGIE S., HAYS J., in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, p. 5007-5015

LI W., SADIGH D., SASTRY S.S., SESHIA S.A., *Synthesis for Human-in-the-Loop Control Systems*, in ÁBRAHÁM E., HAVELUND K. (A CURA DI), *Tools and Algorithms for the Construction and Analysis of Systems*, Berlino-Heidelberg, 2014, p. 470-484

LI X. ET AL., *Deep learning prediction of likelihood of ICU admission and mortality in COVID-19 patients using clinical variables*, in *PeerJ*, 8, 2020, https://peerj.com/articles/10337

LIBERTINI M., *Autorità indipendenti, mercati e regole*, in *Rivista italiana per le scienze giuridiche*, 1, 2010, p. 63 ss.

LIEBNIZ G.W., *Scritti filosofici*, a cura di D. O. BIANCA, Torino, 1978.

LIEW C., QUAH J., GOH H.L., VENKATARAMAN N., *A chest radiography-based artificial intelligence deep-learning model to predict severe Covid-19 patient outcomes: the CAPE (Covid-19 AI Predictive Engine) model*, in *MedRxiv*, 2020, https://doi.org/10.1101/2020.05.25.20113084

LIGHTBOURNE J., *Damned Lies & Criminal Sentencing Using Evidence-Based Tools*, in *Duke Law & Technology Review*, 15, 2017, p. 327 ss.

LIGHTHILL J., *Artificial intelligence: a general survey by Sir James Lighthill*, *FRS Lucasian Professor of Applied Mathematics*, Cambridge, 1972

LILKOV D., *RegulatingArtificial Intelligence in the EU: A Risky Game*, in *EuropeanView*, 20, 2, 2021, p. 166–174

LIM B.Y., DEY A.K., *Assessing demand for intelligibility in context-aware applications*, in *Proceedings of the 11th international conference on ubiquitous computing*, Orlando, 2009, p. 195–204

LIM M., *History of AI winters*, in *Actuaries Digital*, 2018

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, p. 209-261

LIPTON Z.C., *The mythos of model interpretability*, https://doi.org/10.48550/arXiv.1606.03490, 2017.

LLANSÒ E., VON HOBOKEN J., LEERSSEN P., HARAMBAM J., *Artificial intelligence, content moderation and freedom of expression*, Working Paper – Transatlantic Working Group on content moderation online and freedom of expression, 2019

LOENEN T., RODRIGUES P.R. (A CURA DI), *Non-discriminationlaw: comparative perspectives*, Utrecht, 1999

LOEVY R.D., *The Civil Rights Act of 1964: The Passage of the Law That Ended Racial Segregation*, Albany, 1997

LOFTUS T.J. ET AL., *Artificial Intelligence and Surgical Decision-Making*, in *JAMA Surgery*, 155, 2, p. 148 ss.

LOMBARDI G., *Contributo allo studio dei doveri costituzionali*, Milano, 1967

LONDON A.J., *Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability*, in *Hastings Center Report*, 49, 1, 2019, p. 15-21

LONGO G., *Facebook chiude i gruppi Casa Pound e Forza Nuova*, in *La Stampa*, 10 settembre 2019

LOUCKS J., STEWART D., BUCAILLE A., CROSSAN G., *Wearable technology in health care: getting better all the time*, Deloitte, 1 dicembre 2021

LOUW T.L., MADIGAN R., CARSTEN O., MERAT N., *Were they in the loop during automated driving? Links between visual attention and crash potential*, in *Injury Prevention*, 23, 4, 2017

LOUW T.L., MERAT N., JAMSON A.H., *Engaging with highly automated driving: To be or not to be in the loop?*, in *8th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Leeds-Salt Lake City, 2015, https://eprints.whiterose.ac.uk/84892/

LOVELOCK J.,*Novacene: the coming age of hyperintelligence*, Londra, 2019

LU J., *Will Medical Technology Deskill Doctors?*, in *International Education Studies*, 9, 7, 2016, p. 130–134

LU L., ZHANG J., XIE Y., GAO F., XU S., WU X. ET AL., *Wearable health devices in health care: narrative systematic review*, in *JMIR mHealth and uHealth*, 8, 11, 2020, https://doi.org/10.2196/18907

LUCIANI M., *I principi di eguaglianza e non discriminazione: una prospettiva di diritto comparato*, EPRS-Servizio Ricerca del Parlamento Europeo, 2020

LUCIANI M., *La decisionegiudiziariarobotica*, in *Rivista AIC*, 3, 2018, p. 872-893

LUNDBERG S.M., LEE S.I., *A unified approach to interpreting model predictions*, in *Advances in Neural Information Processing Systems - Neural Information Processing Systems Foundatio*n, Long Beach 4–9 December 2017, p. 4765–4774

LUO X. ET AL., *Machines vs humans: the impact of artificial intelligence chatbot disclosure on customer purchases*, in *Marketing science*, 6, 2019, p. 937 ss.

LUX D., *Facebook's hate speech policies censor marginalized users*, Wired, 14 agosto 2017

LUXTOND.D,,*Should Watson Be Consulted for a Second Opinion?*, in*AMA Journal of Ethics*, 2, 2019, p. 131-138

LYCAN W.C., *Explanation and epistemology*, in MOSER P. (A CURA DI), *The Oxford Handbook of epistemology*, Oxford, 2002, p. 408-433

LYNSKEY O., *Control over Personal Data in a Digital Age: Google Spain v AEPD and Mario Costeja Gonzalez*, in *The modern law review*, 78, 3, 2015, p. 522-534

LYONS K., *Trump sues to reinstate his Twitter account*, The Verge, 2 ottobre 2021

MA A., China's controversial social credit system isn't just about punishing people — here's what you can do to get rewards, from special discounts to better hotel rooms, in Business Insider, 3 febbraio 2019

MAC NAMEE B., CUNNINGHAM P., BYRNE S., CORRIGAN O.I., *The problem of bias in training data in regression problems in medical decision support*, in *Artificial Intelligence in Medicine*, 24, 1, 2002, p. 51-70

MACERA L., *Memorie associative e reti di Hopfield*, in *MC microcomputer*, 105, 1991, p. 282-285

MACIAS E., BOQUET G., SERRANO J., VICARIO J.L., IBEAS J., MORELL A., *Novel Imputing Method for the Early Prediction of Sepsis in ICU Using Deep Learning Techniques*, Computing in Cardiology, 2019

MACIOCE F., *L'identità personale in Cassazione: un punto d'arrivo e un punto di partenza* e DOGLIOTTI M., *Il diritto all'identità personale approda in Cassazione*, in *Giustizia civile*, 1, 1985, p. 3049 ss.

MACKENZIE F.C., *Fear the Reaper: how content moderation rules are enforced on social media*, in *International Review of Law, Computers & Technology*, 34, 2, 2020, 128 ss.

MACMILLAN A., *Hospitals Overwhelmed by FluPatients Are TreatingThem in Tents*, Time, 18 gennaio 2018

MAGISTRONI M., *Intelligenza artificiale: un algoritmo predice quanto tempo resta ai malati terminali*, Wired, 22 gennaio 2018

MAGNANI C., *Libertà d'informazione online e fake news: vera emergenza? Appunti sul contrasto alla disinformazione tra legislatori statali e politiche europee*, in *Forum di Quaderni costituzionali – Rassegna*, 4, 2019, p. 16 ss.

MAGRASSI P., BERG T., *A World of Smart Objects,* Gartner research report R-17-2243, 12 agosto 2002

MAGUOLO G., NANNI L., *A critic evaluation of methods for COVID-19 automatic detection from X-ray images*, in*Information Fusion*, 76, 2021, p. 1-7.

MAKASI T., NILI A., ALIREZA T., DESOUZA K., TATE M., *Chatbot-mediated public service delivery: a public service value-based framework*, in *First Monday*, 25, 12, https://eprints.qut.edu.au/204999/

MALDOFF G., *The risk-based approach in the GDPR: interpretation and implications*, IAPP – White Paper, 2016

MALGIERI G., *Automated decision-making in the EU Member States: the right to explanation and other "suitable safeguards" in the national legislations*, in *Computer law & security review*, 25, 2019

MAMANDIPOOR B., MAJD M., MOZ M., OSMANI V., *Blood Lactate Concentration Prediction in Critical Care*, in *Digital Personalized Health and Medicine*, 2020, doi:10.3233/SHTI200125

MAMANDIPOOR B., FRUTOS-VIVAR F., PEÑUELAS O., REZAR R. ET AL., *Machine learning predicts mortality based on analysis of ventilation parameters of critically ill patients: multi-centre validation*, in *BMC Medical Informatics and Decision Making Journal*, 2021

MAMANDIPOOR B., YEUNG W., AGHA-MIR-SALIM L. ET AL., *Prediction of blood lactate values in critically ill patients: a retrospective multi-center cohort study*, in *Journal of Clinical Monitoring and Computing*, 36, 2022, p. 1087–1097, https://doi.org/10.1007/s10877-021-00739-4

MANES V., *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *Discrimen*, 15 febbraio 2020, p. 14-17

MANGIONE C., BOZZI S., *Storia della logica*, Catania, 1985

MANHART K., *Artificial Intelligence Modelling: Data Driven and Theory Driven Approaches*, in TROITZSCH K.G. (a cura di),*Social Science Microsimulation*, Berlino, 1996, p. 416-431

MANHEIM K., KAPLAN L., *Artificial intelligence: risks to privacy and democracy*, in *Yale Journal of Law & Technology*, 21, 2019, p. 106-188

MANN R. J., BELZLEY S. R., *The promise of internet intermediary liability*, in *William and Mary Law Review*, 47, 1, 2005, p. 239-308

MANTELERO A., *Artificial Intelligence and Data Protection: Challenges and Possible Remedies – Report for the Council of Europe consultative committee of the Convention 108* (T-PD(2018)09Rev), 2018

MANTOVANI R., *L'Intelligenza Artificiale predice morti premature meglio dell'uomo*, Sky TG24, 28 marzo 2019

MANTOVANI R., *Ricordati che devi morire? Ci pensa Google*, Focus, 25 giugno 2018

Manuale interistituzionale di convenzioni redazionali, Ufficio delle Pubblicazioni dell'Unione Europea, 2022

MARCHELLO E., *I processi decisionali*, Milano, 2003

MARCHESI A., *La protezione internazionale dei diritti umani*, Torino, 2021

MARCHETTI B., *La garanzia dello "human in the loop" alla prova della decisione amministrativa algoritmica*, in *BioLaw Journal - Rivista di BioDiritto*, 2021, 2, pp. 367-385

MARCHETTI G., *Le fake news e il ruolo degli algoritmi*, in *MediaLaws–Rivista di diritto dei media*, 1, 2020, p. 29-35

MARGOLIS J., *A Big Brother approach has qualities that would benefit society*, Financial Times, 31 ottobre 2017.

MARINI G., *La giuridificazione della persona. Ideologie e tecniche dei diritti della personalità*, in *Rivista di diritto civile*, I, 2006, p. 359 ss.

MARKOFF J., *Computer wins on "Jeopardy!": trivial' It's not*, The New York Times, 17 febbraio 2011

MARKOFF J., *In a grueling desert race, a winner, but not a driver*, The New York Times, 9 ottobre 2005.

MARR B., *Man vs. machine: the 6 greatest AI challenge to showcase thepower of artificial intelligence*, in *Forbes (online)*, 2019

MARRONE M., *Rights against the machine! Food delivery, piattaforme digitali e sindacalismo informale*, in *Labour&Law Issues*, 5, 1, 2019

MARSHALL H., DRIESCHOVA A., *Post-truth politics in the UK's Brexit referendum*, in *New Perspectives*, 26, 3, 2018, p. 89–106

MARTIN K., *Algorithmic Bias and Corporate Responsibility: How companies hide behind the false veil of the technological imperative*, in MARTIN K. (A CURA DI), *Ethics of data and analytics*, Boca Raton, 2022.

MARTIN P., *Why you should not worry about AI black box*, 23 ottobre 2017

MARX K., *Il Capitale*, a cura di M. L. BOGGERI, Roma, 2016

MATHIAS J.N., *Bias and Noise: Daniel Kahneman on Errors in Decision-Making*, Medium, 17 ottobre 2017

MATHOTARACHI S., ZHU M., XU C., YU J., WU Y., LI C., ZHANG M., *Differentiation of Pancreatic Cancer and ChronicPancreatitis Using Computer-AidedDiagnosis of EndoscopicUltrasound (EUS) Images: A Diagnostic Test*, in *PLoS ONE*, 8, 5, 2013

MATSAKIS L., *How the West got China's social credit system wrong*, in *Wired*, 29 luglio 2019

MATTEUCCI N., *Organizzazione del potere e libertà. Storia del costituzionalismo moderno*, Torino, 1976

MATTEUCCI N., *Breve storia del costituzionalismo*, Brescia, 2010

MAUGERI G., COMANDÉ G., *Why a right to legibility of automated decision making exists in the General Data Protection Regulation*, in *International Data Privacy Law*, 7, 4, 2017, p. 243-265

MAZIOTTI DI CELSO M., *Lezioni di diritto costituzionale*, II, Milano, 1985

MCALLESTER D.A., *What is the most pressing issue facing AI and the AAAI today?*,Candidate statement, election for Councilor of the American Association for Artificial Intelligence, 1998

MCCARTHY J., MINSKY M.L., ROCHESTER N., SHANNON C.E., *A proposal for the DartmoutSummerResearch Project on Artificial Intelligence*, 1955

MCCARTHY J., *Recursive functions of symbolic expressions and their computation by machine*, in *Communications of the ACM*, April 1960

MCCARTHY J., *Concept of logical AI*, in MINKER J. (a cura di), *Logic-basedartificial intelligence*, Norwell, 2000, p. 37-56

MCCARTHY J., *What is artificial intelligence?*, Stanford, 2007, http://jmc.stanford.edu/articles/whatisai.html

MCCLELLAND J.L., BOTVINICK M., *Deep learning: Implications for human learning and memory,* in *The Oxford Handbook of Human Memory*, 2020, doi.10.31234/osf.io/3m5sb

MCCORDUCK P., *Machines Who Think: a personal inquiry into the history and prospects of artificial intelligence*, Natick, 2004

MCCULLOCH W.S., PITTS W., *A logical calculus of ideas immanent in nervous activity*, in *Bulletin of Mathematical Biophysics*, 5, 1943, p. 115-133

MCDANIEL J.L.M., PEASE K.G., *Predictivepolicing and artificial intelligence*, 2021, New York

MCDERMID J.A., JIA Y., PORTER Z., HABLI I., *Artificial intelligence explainability: the technical and ethical dimensions*, in *Philosophical Transaction Royal Society A*, 2021, doi:379(2207):20200363

MCDERMOTT J., *R1: a rule-based configurer of computer systems*, in *Artificial Intelligence*, 19, 1, 1982, p. 39-88

MCDERMOTT D., MITCHELL WALDROP M., SCHANK R., CHANDRASEKARAN B, MCDERMOTT J., *The dark ages of AI: a panel discussionat AAAI-84*, in *AI Magazine*, 6, 3, 1985, p. 122-134

MCDONALD A.M., CRANOR L.F., *The Cost of Reading Privacy Policies*, in *I/S: A Journal of Law and Policy for the Information Society*, 2008, p. 543-564

MCENTIRE R., SZALKOWSKI D., BUTLER J., KUO M.S., CHANG M. ET AL, *Application of an automated natural language processing (NLP) workflow to enable federated search of external biomedical content in drug discovery and development*, in *Drug Discovery Today*, 21, 5, 2016, p. 826-835

MCFARLAND M., *ElonMusk: with artificial intelligence we are summoning to the demon*, The New York Times, 24 ottobre 2014

MCGEREVAN W., *Privacy and data protection law*, St. Paul, 2016.

MCINTIRE J.P., MCINTIRE L.K., HAVIG P.R., *Methods for chatbot detection in distributed text-based communications*, in *International Symposium on Collaborative Technologies and Systems*, 2010, p. 463 ss.

Medicina difensiva. Ci costa 10 mld l'anno. La pratica almeno una volta al mese quasi l'80% dei medici. Il report del Ministero della Salute, in *Quotidiano sanità*, 26 marzo 2015

MEIJER A., WESSELS M., *Predictive policing: review of benefits and drawbacks*, in *International Journal of Public Administration*, 42, 12, 2019, p. 1031 ss.

MELE C., RUSSO SPENA T., KAARTEMO V., MARZULLO M.L., *Smart nudging: how cognitive technologies enable choice architectures for value co-creation,* in *Journal for business research*, 129, 2021, p. 946-960

MELZI D'ERIL C., *Fake news e responsabilità: paradigmi classici e tendenze incriminatrici*, in *MediaLaws – Rivista di diritto dei media*, 1, 2017, p. 60 ss

MELZI D'ERIL C., VIGEVANI G.E., *Odio in rete e rimozione delle pagine Facebook: giudice che vai, soluzione che trovi*, in *Il Sole 24 Ore*, 27 febbraio 2020

MENDELSON W., *Clear and Present Danger--From Schenck to Dennis*, in *Columbia Law Review* 52, 3, 1952, p. 312 ss.

MENECEUR Y., *L'intelligence artificielle en procès. Plaidoyer pour une réglementationinternationale et européenne*, Bruxelles, 2020, p. 96-97

MENEGHETTI F., ROSSI CHAUVENET C., FIORONI G., *Rapporto 3/2022 – SMART cities e intelligenza artificiale*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2022, p. 253 ss.

MENEZES A.J., VON OORSCHOT P.C., VANSTONE S.A., *Handbook of applied cryptography*, 1997

MERAT N. ET AL., *The "Out-of-the-Loop" concept in automated driving: proposed definition, measures and implications*, in *Cognition, Technology & Work*, 21, 1, 2019

MERZ R., O'SULLIVAN D., *A deepfake video of Mark Zuckerberg presents a new challenge for Facebook*, CNN Business, 12 giugno 2019

MESKE C., BUNDE E., SCHNEIDER J., GERSCH M., *Explainable artificial intelligence: objectives, stakeholders, and future research opportunities*, in *Information System Management*, 39, 1, 2020, p. 53-63

MESSER N., *Healthcare Resource Allocation and the 'Recovery of Virtue'*, in *Studies in Christian Ethics*, 18, 1, 2005

MESSINETTI D., *Personalità (diritti della)*, in *Enciclopedia del diritto*, XXXIII, 1983, p. 355

MEZZETTI L., *Diritti umani: protezione internazionale e ordinamenti nazionali*, Pisa, 2021

MILLER S.J., HICKSON D.J., WILSON D.C., *Decision-making in organizations*, in CLEGG S.R., HARDY C., NORD W.R. (A CURA DI), *Managing organizations. Currentissues*, 1999, p. 43-63

MILLER T., *Explanation in artificial intelligence: insights from the social sciences*, in *Artificial intelligence*, 267, 2019, p. 1-38

MILLI S.,SCHMIDT L.,DRAGAN A., HARDT M., *Model Reconstruction from Model Explanations*, http://arxiv.org/abs/1807.05185

MILLS S., *Into hyperspace: an analysis of hypernudges and personalized behavioural science*, 2019, https://doi.org/10.2139/ssrn.3420211

MINH D.,WANG X., LI Y., NGUYEN N.T., *Explainable artificial intelligence: a comprehensive review*, in *Artificial Intelligence Review*, 55, 5, 2022, p. 3503-3568

MINSKY M., PAPERT S., *Perceptrons: an introduction to computational geometry*, Cambridge (US), 1969

MIRACOLA S., *How China usesartificial intelligence to control society*, 3 giugno 2019

MISHKIN P.J., *The uses of ambivalence: reflections of the Supreme Court and the constitutionality of affirmative action*, in *University of Pennsylvania Law Review*, 131, 1983, p. 107 ss.

MISTREANU S., China is implementing a massive plan to rank its citizens, and many of them want in, in Foreign Policy, 3 aprile 2018

MISTREANU S., *Life Inside China's Social Credit Laboratory*, in *Foreign Policy*, 3 aprile 2018

MITCHAM C., *Thinking through technology*. *The path between engineering and philosophy*, Chicago, 1994

MITROU L., *Data protection, artificial intelligence and cognitive services. Is the General Data Protection Regulation (GDPR) Artificial Intelligence-proof?*, 2018, https://dx.doi.org/10.2139/ssrn.3386914

MODUGNO F., *I nuovi diritti nella giurisprudenza costituzionale*, Torino, 1995

MOLNÁR-GÁBOR F., *Artificial Intelligence in Healthcare: Doctors, Patients and Liabilities*, in WISCHMEYER T., RADEMACHER T. (A CURA DI), *RegulatingArtificial Intelligence*, Cham, 2020, https://doi.org/10.1007/978-3-030-32361-5_15

MOLNAR-GABOR F., *Data Protection*, in *Max Planck Encyclopedia of Comparative Constitutional Law*, Oxford, 2016

MONETT D., LEWIS C.W.P., *Getting clarity by defining Artificial Intelligence - A Survey*, in V.C. MULLER (ED.), *Philosophy and Theory of Artificial Intelligence*, Berlino, 2017

MONREALE A., *Rischi etico-legali dell'intelligenza artificiale*, in *DPCE Online*, 3, 2020, p. 3391-3398

MONTANARI L., *I diritti dell'uomo nell'area europea tra fonti internazionali e fonti interne*, Torino, 2002

MONTAVON G., SAMEK W., MÜLLER K.R., *Methods for interpreting and under standing deep neural networks*, in *Digital Signal Processing*, 73, 2017, p. 1-15

MONTI M., *La Corte di giustizia, la direttiva e-commerce e il controllo contenutistico online: le implicazioni della decisione C 18-18 sul discorso pubblico online e sul ruolo di Facebook*, in *MediaLaws*, 3, 2019, p. 1 ss.

MONTI M., *Privatizzazione della censura e internet platforms: la libertà di espressione e i nuovi censori dell'agorà digitale*, in *Rivista italiana di informatica e diritto*, 1, 2019, p. 35 ss.

MONTI M., *La disinformazione online, la crisi del rapporto pubblico-esperti e il rischio della privatizzazione della censura*, in *Federalismi.it*, 11, 2020

MONTI M., *La Corte Suprema statunitense e il potere delle piattaforme digitali: considerazioni sulla privatizzazione della censura a partire da una concurring opinion*, in *DPCE online*, 1, 2021

MORAVEC H., *The great 1980s AI bubble: a review of the brain makers*, in *AI Magazine*, 15, 3, 1994, p. 86-87

MORELLI A., *Persona e identità personale*, in *BioLaw Journal – Rivista di BioDiritto*, Special Issue 2, 2019, p. 45 ss.

MORI M., *The uncanny valley*, in *IEEE Robotics & Automation Magazine*, 2, 2012, p. 98 ss. (originale *Bukimi no tani*, in *Energy*, 4, 1970, 33 ss.)

MORRONE A., *Il bilanciamento nello stato costituzionale: teoria e prassi delle tecniche di giudizio nei conflitti tra diritti e interessi costituzionali*, Torino, 2014

MORTATI C., *La Corte costituzionale e i presupposti della sua vitalità*, in *Iustitia*, 1949, p. 69 ss.

MOSIER K.L., SKITKA L.J., *Human Decision Makers and Automated Decision Aids: Made for Each Other?*, in *Automation and Human Performance: Theory and Applications*, Boca Raton, 1996

MOSTERT M., BREDENOORD A.L., BIESAART M.C.I.H., VAN DELDEN J.J.M., *Big Data in medicalresearch and EU data protectionlaw: challenges to the consent or anonymiseapproach*, in *European Journal of Human Genetics*, 24, 7, 2016, p. 956 ss.

Mostpopular social networks worldwideas of January 2022, ranked by number of monthlyactive users, Statista.com, https://bit.ly/3GXtGb0

MOURON P., *Une future loi pour luttercontrelesfake news: lesdifficultés d'une définitionjuridique*, in *Revue européenne desmédias et dunumérique*, 2018, 45, p. 66 ss.

MUCIACCIA N., *Algoritmi e procedimento decisionale: alcuni recenti arresti della giustizia amministrativa*, in *Federalismi.it*, 15 aprile 2020

MUELLER B., *How Much Will the Artificial Intelligence Act Cost Europe?*, Report - Center for Data Innovation, 26 luglio 2021

MURRAY R., *The African Charter of Human and Peoples' Rights. A commentary*, Oxford, 2019

MUSSELLI L., *La decisione amministrativa nell'età degli algoritmi: primi spunti*, in *MediaLaws – Rivista di diritto dei media*, 1, 2020, p. 18-28

NAVARETTA E., *Principio di uguaglianza, principio di non discriminazione e contratto*, in *Rivista di diritto civile*, 2014, p. 547-566

NEWELL A., SHAW J.C., SIMON H.A., *Report on a general problem-solving program*, 1959, https://bit.ly/2GmYirj

NEWELL A., SIMON H., *The logic theory machine: a complex information processing system*, RAND Corporation - report, 15 giugno 1956

NEWQUIST H.P., *The brain makers: genius, ego, and greed in the quest for machines that think*, Indianapolis, 1994

NGUYEN P., TRAN T., VENKATESH S., *Deep Learning to Attend to Risk in ICU*, 2017, https://arxiv.org/abs/1707.05010

NICHOLSON PRICE II W., GERKE S., COHEN I.G., *Potential Liability for Physicians Using Artificial Intelligence*, in *JAMA*, 322, 18, 2019, p. 1751 ss.

NICITA A., *Le piattaforme online tra moderazione e autoregolazione: verso il Digital Services Act*, in *MediaLaws*, 25 novembre 2020

NICOTRA I., VARONE V., *L'algoritmo, intelligente ma non troppo*, in *Rivista AIC*, 4, 2019, p. 86 ss.

NIILER E., *Can AI Be a Fair Judge in Court? Estonia Thinks So*, Wired, 25 marzo 2019

NILSSON N.J., *Artificial intelligence: a new synthesis*, Burlington, 1998

NILSSON N.J., *Logic and artificial intelligence*, in *Artificial Intelligence*, 47, 1991, p. 31-46

NILSSON N.J., *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, Cambridge, 2009

NIRALA K., SINGH K.N., PURANI V.S., *A survey on providing customer and public administration-based services using AI: chatbot*, in *Multimed Tools Appl* 81, 2022, p. 22215–22246

NOALE M., LIMONGI F., SCAFATO E., MAGGI S., CREPALDI G., *Longevity and health expectancy in an ageing society: implications for public health in Italy*, in *Annalidell'IstitutoSuperiore di Sanità*, 48, 3, 2012, p. 292-299

NOSSEM E., *Queer, frocia, femminiellə, ricchione et al. Localising "queer" in the Italian context*, in *Gender/Sexuality/Italy*, 6, 2019

NOUVEAU P., *European Union's digital governance vs. United States' digital dominance*, in *Revue de la Faculté de droit de l'Université de Liège*, 2, 2020, p. 208-232

NTOUTSI E. ET AL., *Bias in Data-driven Artificial Intelligence Systems—An Introductory Survey*, in *WIREs Data Mining and Knowledge Discovery*, 10, 3, 2020

NUCKOLLS C.W., *The Anthropology of Explanation*, in *Anthropological Quarterly*, 66, 1, 1993, p. 1 ss.

NUNES D.S., ZHANG P., SÁ SILVA J., *A Survey on Human-in-the-Loop Applications Towards an Internet of All*, in *IEEE Communications Surveys & Tutorials*, 17, 2, 2015, p. 944-965

O'NEILL C., *Weapons of math destruction: how big data increases inequality and threatens democracy*, Largo (USA), 2016

OCSE CONSIGLIO DEI MINISTRI, *Recommendation on artificial intelligence*, 22 maggio 2019, OECD/LEGAL/0449

ODENNINO A., *Decisioni algoritmiche e prospettive internazionali di valorizzazione dell'intervento umano*, in *DPCE online*, 1, 2020, p. 199 ss

OFFICE FOR THE PRIVACY COMMISSIONER OF CANADA, *Consent and privacy. A discussion paper exploring potential enhancements to consent under the Personal Information Protection and Electronic Documents Act*, 2016, https://www.priv.gc.ca/media/1806/consent_201605_e.pdf

O'LEARY D.E., *Artificial Intelligence and Big Data*, in *IEEE Intelligent Systems*, 28, 2, 2013, p. 96-99

OLIVETTI M., *Diritti fondamentali*, Torino, 2020, p. 3

OLSEN S., *Spying an intelligent search engine*, in *www.cnet.com*, 2006, https;//cnet.co/2Z4ccVN

OLVER M.E., *Some considerations on the use of actuarial and related forensic measures with diverse correctional populations*, in *Journal of Threat Assessment and Management,* 3, 2, 2016, p. 107–121

OMERO, *Iliade*, a cura di V. MONTI, Milano, 1810

One LEGALE presenta la nuova funzionalità Giurimetria, Studio Cataldi – il diritto quotidiano, 8 settembre 2021, https://bit.ly/3BRDe6M (19 agosto 2022).

ORDONSELLI N., *"Porno deepfake": profili di diritto penale*, in *CyberLaws*, 18 gennaio 2021

ORFALI K., *Getting to the Truth: Ethics, Trust, and Triage in the United States versus Europe during the Covid-19 Pandemic*, in *Hastings Center Report*, 51, 1, 2021

ORFALI K., *What Triage Issues Reveal: Ethics in the COVID-19 Pandemic in Italy and France*, in *Journal of Bioethical Inquiry*, 17, 4, 2020

OROFINO M., *La libertà di espressione tra Costituzione e Carte europee dei diritti*, Torino, 2014

ORSEAU L., ARMSTRONG M., *Safely interruptible agents*, in *Conference on Uncertainty in Artificial Intelligence*, 2016

ORTEGA Y GASSET J., *Meditación de la técnica*, in *Ensimismamiento y alteración. Meditación de la técnica*, Madrid, 1939

OSUNWUSI A., *Aviation Automation and CNS/ATM-related Human-Technology Interface: ATSEP Competency Considerations*, in *International Journal of Aviation, Aeronautics, and Aerospace*, 6, 4, 2019

OSWALD M., GRACE J., URWIN S., BARNES G.C., *Algorithmic risk assessmentpolicing models: lessons from the Durham HART model and "Experimental" proportionality*, in *Information and Communications Technology Law*, 2018, p. 227 ss.

OVIDE S., *How big tech won the pandemic*, The New York Times, 30 aprile 2021

ÖZTÜRK H., ÖZGÜR A., SCHWALLER P., LAINO T., OZKIRIMLI E., *Exploring chemical space using natural language processing methodologies for drug discovery*, in *Drug Discovery Today*, 25, 4, 2020, p. 689-705

PACE A., MANETTI M., *La libertà di manifestazione del pensiero. Art. 21*, in G. BRANCA, A. PIZZORUSSO, *Commentario della Costituzione*, XI, Bologna, 2006

PACE A., *Problematica delle libertà costituzionali. Parte generale*, 2003

PAEZ A., *The Pragmatic Turn in Explainable Artificial Intelligence (XAI)*, in *Minds and Machines*, 29, 2019, p. 441-459

PAGALLO U., QUATTROCOLO S., *The impact of AI inCriminalLaw, and itsTwofoldProcedures*, in BARFIELD W., PAGALLO U. (A CURA DI),*ResearchHandbook on the Law of Artificial Intelligence*, Cheltenam, 2018, 388 ss.

PAIKENS P., ZNOTIŅŠ A., BĀRZDIŅŠ G., *Human-in-the-loop conversation agent for customer service*, in MÉTAIS E., MEZIANE F., HORACEK H., CIMIANO P. (A CURA DI), *Natural Language Processing and information systems*, Cham, 2020

PAJNO A. E AL., *AI: profili giuridici. Intelligenza artificiale: criticità emergenti e nuove sfide per i giuristi*, in *Biolaw Journal*, 3, 2019, p. 205 ss.

PALADIN L., *Eguaglianza (diritto cost.)*, in *Enciclopedia del diritto*, XIV, Milano, 1965, p. 519 ss.

PALADIN L., *Il principio costituzionale d'eguaglianza*, Milano, 1965

PANAYOTOV V., CHEN G., POVEY D., KHUDANPUR S., *LibriSpeech: an ASR corpus based on public domain audiobooks*, in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, 2015, p. 5206-5210

PANDE V., *Artificial intelligence's "black-box" is nothing to fear*, The New York Times, 25 gennaio 2018

PANTING G., *Doctors on the defensive*, The Guardian, 1 aprile 2005

Para la Corte Suprema, losbuscadores no son responsables del contenidoquelistan, La Nación, 14 gennaio 2015

PARDOLESI R., *(Protezionedeidatipersonali) Nota a sent. CGUE grande sez. 24 settembre 2019 (causa C-507/17 Google France vs CNIL); sent. CGUE grande sez. 24 settembre 2019 (causa C-136/17)*, in *Foro italiano*, 12, 4, 2019, p. 594-597

PARDOLESI R., *Diritti della personalità*, in *AIDA*, 2005, p. 3 ss.

PARDOLESI R., *Le nuove tabelle milanesi e il fascino discreto della para-normatività*, in *Danno e responsabilità*, 26, 4, 2021, p. 423-432

PARISIER E., *The filter bubble: what the internet is hiding from you*, New York, 2011

PARKER D.B., *Learning Logic*, MIT Center for Computational Research in Economics and Management Science – Technical Report, 1985

PARLAMENTO EUROPEO, *Risoluzione recante taccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica* (2015/2103(INL)), 16 febbraio 2017

PARLAMENTO EUROPEO, *Risoluzione recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale*, *della robotica e delle tecnologie correlate*. 2020/2012(INL), 20 ottobre 2020

PARLAMENTO EUROPEO, *Risoluzionerecante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale*, 2020/2014(INL), 20 ottobre 2020

PARLAMENTO EUROPEO, *Risoluzione recante sui diritti di proprietà intellettuale per lo sviluppo di tecnologie di intelligenza artificiale*, 2020/2015(INI), 20 ottobre 2020

PASCOAL T.A., SHIN M., BENEDET A.L., KANG M., BEAUDRY T., *Identifyingincipientdementiaindividualsusing machine learning and amyloid imaging*, in *Neurobiology of Aging*, 59, 2017, p. 80 ss.

PASCUZZI G., DUCATO R., *Biobanche di ricerca tra proprietà, privacy e proprietà intellettuale: un approccio LawTech*, in *Notizie di Politeia*, 2012, p. 55 ss.

PASCUZZI G., *Il diritto dell'era digitale*, Bologna, 2020, p. 47 ss.

PASQUALE F., *The Black-Box Society: The Secret Algorithms That Control Money and Information*, Harvard, 2016

PATEL D. ET AL., *Machine learning basedpredictors for COVID-19 disease severity*, in *Scientific Reports*, 11, 2021, p. 4673 ss.

PAWLICKA A., PAWLICKI M., KOZIK R., CHORAŚ M., *Human-driven and human-centred cybersecurity: policy-making implications,* in *Transforming Government: People, Process and Policy*, 2022 (in corso di pubblicazione), disponibile in: https://doi.org/10.1108/TG-05-2022-0073

PEARL J., *Probabilistic reasoning in intelligent systems*, Burlington, 1986

PECES-BARBA MARTÍNEZ G., *Curso de derechosfondamentales. Teoría general*, Madrid, 1991

PEDRESCHI D., GIANOTTI F., GUIDOTTI R., MONREALE A., RUGGIERI S., TURINI F., *Meaningful explanations of Black Box AI decision systems*, in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 10.1609/aaai.v33i01.33019780

PEIRÓ J.M., MELIÁ J.L., *Formal and informalinterpersonal power in organisations: testing a bifactorial model of power in role-sets*, in *Applied psicology*, 52, 1, 2003, p. 14-35

PELLICER S., SANTA G., BLEDA G.A., MAESTRE R., JARA A.J., SKARMETA A.G., *A global perspective of smart cities: a survey*, in *Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, 2013, p. 439-444

PENASA S., *Intelligenza artificiale e giustizia: il delicato equilibrio tra affidabilità tecnologica e sostenibilità costituzionale in prospettiva comparata*, in ***DPCE Online***, 1, 2022, p. 297-310

PENNISI M., *CasaPound e Forza Nuova rimossi definitivamente da Facebook e Instagram: «Diffondono odio»*, in *Corriere della Sera (online)*, 9 settembre 2019

PEREGO B., *Predictivepolicing: trasparenza degli algoritmi, impatto sulla privacy e risvolti discriminatori*, in *BioLaw Journal - Rivista di BioDiritto*, 2, 2020, p. 447-465

PERELMAN C., OLBRECHTS-TYTECA L., *Traité de l'argumentation, la nouvelle rhétorique*, Bruxelles, 1958

PERETTI T., *Constructing the State Action Doctrine, 1940–1990*, in *Law & Social Inquiry*, 35, 2, 2010, 273 ss.

PERIN A., *Standardizzazione, automazione e responsabilità medica. Dalle recenti riforme alla definizione di un modello d'imputazione solidaristico e liberale*, in *BioLaw Journal - Rivista di BioDiritto*, 1, 2019, p. 29 ss.

PERRONE R., *Fake news e libertà di manifestazione del pensiero: brevi coordinate in tema di tutela costituzionale del falso*, in *Nomos-Le attualità del diritto*, 2, 2018, p. 25 ss.

PERRY W., MCINNIS B., PRICE C.C., SMITH S.C., HOLLYWOOD J., *Predictive policing. The role of crime forecasting in law enforcement operations*, Washington, 2013

PETERSMANN E.U., *Multilevel constitutionalism for multilevel governance of public goods*, Oxford-Portland, 2017

PETRALIA V., *Equo processo, giudicato nazionale e convenzione europea dei diritti dell'uomo*, Torino, 2012

PHILBECK T., DAVIS N., *The fourth industrial devolution: shaping a new era*, in *Journal of International affairs*, 72, 1, 2018, p. 17 ss.

PHILLIPS-WREN G., JAIN L., *Artificial Intelligence for Decision Making*, in *Knowledge-Based Intelligent Information and Engineering Systems*, Berlino-Heidelberg, 2006, p. 531 ss.

PICCIALLI F., DI SOMMA V., GIAMPAOLO F., CUOMO S., FORTINO G., *A survey on deep learning in medicine: Why, how and when?*, in *Information Fusion*, 66, 2021, p. 111 ss.

PIERSON J., HEYMAN R., *Social media and cookies: challenges for online privacy*, in *Info*, 13, 6, 2011, p. 30 ss.

PINESCHI L., *Diritti umani (protezione internazionale dei)*, in *Enciclopedia del diritto – Annali*, V, 2012

PINO G., *Il diritto all'identità personale. Interpretazione costituzionale e creatività giurisprudenziale*, Bologna, 2003

PINO G., *Teoria e pratica del bilanciamento: tra libertà di manifestazione del pensiero e tutela dell'identità personale*, in *Danno e responsabilità*, 6, 2003, p. 577 ss.

PINO G., *Teorie e dottrine dei diritti della personalità. Uno studio di meta-giurisprudenza analitica*, in *Materiali per una storia della cultura giuridica*, 1, 2003, p. 237 ss.

PINO G., *Il diritto all'identità personale ieri e oggi. Informazioni, mercato, dati personali*, in PANETTA R., *Libera circolazione e protezione dei dati personali*, Milano, 2006, p. 257 ss.

PINO G., *Diritti e interpretazione*, Bologna, 2010

PINO G., *Identità personale*, in RODOTÀ S., TALLACCHINI M. (a cura di), *Ambito e fonti del biodiritto*, in RODOTÀ S., ZATTI P. (diretto da),*Trattato di biodiritto*, Milano, 2010, p. 297 ss.

PINO G.,*Il costituzionalismo dei diritti*, Bologna, 2017

PITRUZZELLA G., POLLICINO O., QUINTARELLI S., *Parole e potere. Libertàd'espressione, hate speech e fake news*, Milano, 2017

PITT J.C. (A CURA DI), *Theories of explanation*, Oxford, 1988

PIVA S., *Facebook è un servizio pubblico? La controversia su CasaPound risolleva la "quaestio" dell'inquadramento giuridico dei "social network"*, in *dirittifondamentali.it*, 2, 2020, p. 1192 ss.

PIZZETTI F.G., *Aspetti e problemi del costituzionalismo multilivello*, Milano, 2004

PIZZORUSSO A., *Che cos'è l'eguaglianza, il principio etico e la norma giuridica nella vita reale*, Roma, 1983

POLIDORO D., *Tecnologie Informatiche e Procedimento Penale: La Giustizia Penale 'Messa Alla Prova' Dall'intelligenza Artificiale*, in *Archivio penale*, 3, 2020

POLLICINO O., BASSINI M., *Free speech, defamation and the limits to freedom of expression in the EU: a comparative analysis*, in SAVIN A., TRZASKOWSKI J. (A CURA DI), *ResearchHandbook On EU Internet Law*, Cheltenham-Northampton, 2014, p. 508 ss.

POLLICINO O., *Tutela del pluralismo nell'era digitale: ruolo e responsabilità degli Internet service provider*, in *Percorsi costituzionali*, 2014, 1, p. 45 ss.

POLLICINO O., *L'efficacia orizzontale dei diritti fondamentali previsti dalla Carta. La giurisprudenza della Corte di giustizia in materia di digital privacy come osservatorio privilegiato*, in *MediaLaws*, 3, 2018

POLLICINO O., *L' "autunno caldo" della Corte di giustizia in tema di tutela dei diritti fondamentali in rete e le sfide del costituzionalismo alle prese con i nuovi poteri privati in ambito digitale*, in *Federalismo.it*, 19, 2019

POMEROL J.C., *Artificial Intelligence and Human Decision Making*, in *European Journal of Operational Research* 99, 1, 1997, p. 3–25

POPPEL D., IDSARDI W., *We don't know how the brain stores anything, let alone words*, in *Trends in Cognitive Sciences*, 26, 12, 2022, p. 1054-1055

PREMACK D., WOODRUF G., *Does the chimpanzee have a theory of mind*?, in *Behavioral and brain sciences*, 1978, 1, 4, p. 515-526

PRINCE A.E., SCHWARCZ D., *Proxy discrimination in the age of artificial intelligence and big data*, in *Iowa Law Review*, 105, 2019, p. 1257 ss.

Proposition de loi constitutionelle n. 2585 relative à la Charte de l'intelligence artificielle et des algorithmes, 15 gennaio 2020

PROSSER W., *Privacy*, in *California Law Review*, 48, 1960, p. 383 ss.

PRUITI CIARELLO A., *Oggi la task force anti fake news. Domani? L'opinione della Fondazione Einaudi*, in *formiche.net*, 2020

PSEUDO-APOLLODORO, *Bibliotheca*, I, 9, 26

PUBLIO OVIDIO NASONE, *Le Metamorfosi*, X, 243-297

PUGLIESE G., *Una messa a punto della Cassazione sul preteso diritto alla riservatezza*., in *Giur. it.* 1957, I, p. 366 ss.

PURTOVA N., *Property in personal data: a European perspective on the instrumentalist theory of propertization*, in *Law and technology – selected essays*, Pistoia, 2009, p. 225 ss.

PURTOVA N., *Property rights in personal data: learning from the American discourse*, in *Computer law & security review*, 25, 6, 2009, p. 507 ss.

PURTOVA N., *The illusion of personal data as no one's property*, in *Law, innovation & technology*, 7, 1, 2015, p. 83 ss.

QUARTA A., *Disattivazione della pagina Facebook. Il caso CasaPound tra diritto dei contratti e bilanciamento dei diritti*, in *Danno e responsabilità*, 4, 2020, p. 489 ss.

QUATTROCOLO S., *Equo processo penale e sfide della società algoritmica*, in *Rivista di BioDiritto – BioLaw Journal*, 1, 2019, p. 135 ss.

QUATTROCOLO S., *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs. rischi e paure della giustizia digitale "predittiva"*, in *Cassazione penale*, 59, 4, 2019, p. 1748 ss.

QUATTROCOLO S., *Artificial Intelligence, ComputationalModelling and CriminalProceedings. A Framework for A European Legal Discussion*, Berlino, 2020

QUATTROCOLO S., *Forecasting the future whileinvestigating the past. The use of computational models in pre-trial detentiondecisions*, in *Revista Brasileira de Direito Processual Penal*, 7, 3, 2021

QUATTROCOLO S., *Risk assessment: sentencing o non sentencing?*, in A.A.V.V., *Giurisdizione penale, intelligenza artificiale ed etica del giudizio*, Milano, 2021

QUINTARELLI S., *Content moderation: i rimedi tecnici cit., p. 119-120;* G. CHASTEL*, Why is Natural Language Processing still so innatural?*, in *Newton X-access knowledge*, 27 marzo 2018

QUINTARELLI S., *Capitalismo immateriale. Le tecnologie digitali e il nuovo conflitto sociale*, Torino, 2019

QUINTARELLI S., COREA F., FOSSA F., LOREGGIA A., SAPIENZA S., *AI: Profili etici. Una prospettiva etica sull'intelligenza artificiale: principi, diritti e raccomandazioni*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2019, p. 195 ss.

QUINTARELLI S. (A CURA DI), *Intelligenzaartificiale. Cos'è davvero, come funziona, che effetti avrà*, Torino, 2020

RAFFIOTTA E., *Appunti in materia di diritto all'identità personale*, in *www.forumcostituzionale.it*, 26 gennaio 2010

RAFFIOTTA E., *L'erompere dell'intelligenza artificiale per lo sviluppo della pubblica amministrazione e dei servizi al cittadino*, in G.C. FERONI, C. FONTANA, E.C. RAFIOTTA, *AI ANTHOLOGY - Profili giuridici, economici e sociali dell'intelligenza artificiale*, Bologna, 2022

RAFFIOTTA E., *Artificial intelligence, identification tools and identity protection*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2022, p. 165-179.

RAI A., *Explainable AI: from black box to glass box*, in *Journal of the Academy of Marketing Science*, 48, 1, 2020, p. 137 ss.

RAJKOMAR A. ET AL., *Scalable and accurate deep learning with electronic health records*, in *Nature Digital Medicine*, 1, 1, 2018

RAJPURKAR P. ET AL., *CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning*, 2017, arXiv:1711.0522

RANDAZZO B., *Giustizia costituzionale sovranazionale. La Corte europea dei diritti dell'uomo*, Milano, 2012

RAVÌ PINTO R., *Brevi considerazioni su stato di emergenza e diritto costituzionale*, in *BioLaw Journal – Rivista di BioDiritto*, Special Issue 1, 2020, p. 43 ss.

RAVIZZA S.,*Influenza, ospedali in tilt fra tagli e psicosi vaccino: "Lorenzin arrivata tardi"*, il Fatto Quotidiano, 24 gennaio 2015

RAVIZZA S.,*Milano, terapie intensive al collasso per l'influenza: già 48 malati gravi, operazioni rinviate*, Corriere della Sera, 10 gennaio 2018

RAZ J., *Law, morality and authority*, in *The Monist*, 68, 3, 1985, p. 295 ss.

RAZZANO G., *Riflessioni a margine delle raccomandazioni SIAARTI per l'emergenza Covid-19, fra triage, possibili discriminazioni e vecchie DAT: verso una rinnovata sensibilità per il diritto alla vita?*, in *Rivista AIC*, 3, 2020, p. 107 ss.

REALE C.M., TOMASI M., *Libertà d'espressione, nuovi media e intelligenza artificiale: la ricerca di un nuovo equilibrio nell'ecosistema costituzionale*, in *DPCE online*, 51, 1, 2022, p. 331 ss.

REARDON M., *Section 230: how it shields Facebook and why Congress wants changes*, CNET, 6 ottobre 2021

Recent case: Knight First Amendment Institute at Columbia University v. Trump, in *Harvard Law Review Blog,* 3 giugno 2019

REDMAN J., FLETCHER D.R., *Violent bureaucracy: A critical analysis of the British public employment service*, in *Critical Social Policy*, 42, 2, 2022

REPORTERSWITHOUTBORDERS, *2022 World Press Freedom Index*, https://rsf.org/en

REPUBLIC OF ESTONIA – JUSTIITSMINISTEERIUM, *Estonia doesnotdevelop AI Judge*, 16 febbraio 2022

RESCIGNO P., *Il principio di eguaglianza nel diritto privato*, in *Rivista trimestrale dir. proc. civ.*, 1959, p. 1515 ss.

RESCIGNO P., *Sul cosiddetto principio d'uguaglianza nel diritto privato*, in *Foro it.*, 1959, p. 664 ss.

RESCIGNO P., *Personalità (diritti della)*, in *Enciclopedia giuridica*, XXIV, 1990

RESTA G., *Autonomia privata e diritti della personalità*, Napoli, 2005; *Diritti della personalità: problemi e prospettive*, in *Diritto dell'informazione e dell'informatica*, 2007, p. 1043 ss.

RESTA G., ZENO-ZENCOVICH V. (A CURA DI), *Il diritto all'oblio su internet dopo la sentenza Google Spain*, Roma, 2015

RIBEIRO M.T., SINGH S., GUESTRIN C., *"WhyShould I Trust You?": Explaining the Predictions of AnyClassifier*, 2016, http://arxiv.org/abs/1602.04938

RICH E., KNIGHT K., *Artificial intelligence (2 ed.)*, New York, 1991

RICHARDS N.M., SMART W.D., *How should the law think about robots?*, in CALO R., FROOMKIN A.M., KERR I. (a cura di), *Robot Law*, Cheltenham, 2016, p. 3-22

RIFKIN J., *La fine del lavoro. Il declinodella forza lavoroglobale e l'avventodell'era post-mercato*, Milano, 1995

RIMEDIO A., *Criteri di priorità per l'allocazione di risorse sanitarie scarse nel corso della pandemia da CoViD-19*, in *BioLaw Journal - Rivista di BioDiritto*, 1, 2021

RINALDI L., *Le piattaforme tra diritto pubblico e diritto privato: libertà d'espressione, discorso politico e social network in alcuni casi recenti tra Italia e Stati Uniti*, in *Gruppo di Pisa. Dibattito aperto sul Diritto e la Giustizia costituzionale*, Quad. Monografico n. 3 - fascicolo 2, 2021, p. 223 ss.

RINALDIS A., *Una task force contro le false notizie? Il monopolio della verità è l'anticamera del totalitarismo*, in *Globalist.it*, 9 aprile 2020

RINI R., *Deepfakes are coming. We can no longer believe what we see*, The New York Times, 10 giugno 2019

RIOFRÍOMARTÍNEZ VILLALBA J.C., *La cuarta ola de derechoshumanos: losderechosdigitales*, in *Revistalatinoamericana de derechoshumanos*, 25, 1, 2014, p. 15 ss.

RIORDAN J., *The liability of internet intermediaries*, Oxford, 2016

RISSE M., *Human Rights and Artificial Intelligence: An Urgently Needed Agenda*, in *HKS Faculty Research Working Paper Series RWP18-015*, 2018

RIVERA I., *Le tonalità dell'ambiente e le generazioni future nel cammino di riforma della Costituzione*, in *BioLaw Journal - Rivista di BioDiritto*, 2, 2022, p. 225 ss.

ROBERTS H., COWLS J., MORLEY J., TADDEO M., WANG V., FLORIDI L., *The Chinese approach to artificial intelligence: an analysis of policy, ethics and regulation*, in *AI & Society*, 2021, p. 61 ss.

ROBERTS J., *Trump, Twitter, and the First Amendment*, in *Alternative Law Journal*, 44, 3/2019, 207 ss.

ROBERTS S.T., *Behind the screen: the hidden digital labor of commercial content moderation*, 2014, http://hdl.handle.net/2142/50401

ROBERTSON A., *Go readabouthowFacebook's pseudo-Supreme Court come together*, The Verge, 12 febbraio 2021

ROBERTSON G., *Is artificial intelligence (AI) just a buzzword?* in *Speechmatics*, 2019, https://bit.ly/2QETXBJ

ROBNIK-SIKONJA M., KONONENKO I., *Explaining Classifications For Individual Instances*, in *IEEE Transactions on Knowledge and Data Engineering*, 20, 5, 2008, p. 589 ss.

ROCHE J.P., *Equality in America: the expansion of a concept*, in *North Carolina Law Review*, 43, 2, 2, 1969, p. 249 ss.

RODOTÀ S., *Elaboratori elettronici e controllo sociale*, Bologna, 1973

RODOTÀ S., *Tecnologie e diritti*, Bologna, 1995

RODOTÀ S., *Privacy, libertà, dignità*, Discorso conclusivo della XVI Conferenza internazionale sulla protezione dei dati, Wroclaw (PL), 14, 15, 16 settembre 2004

RODOTÀ S., *Intervista su privacy e libertà* (A CURA DI P. CONTI), Bari, 2005

RODOTÀ S., ZATTI P. (DIRETTO DA), *Trattato di biodiritto*, Milano, 2010

RODRIGUEZ D.B., WEINGAST W.B., *The Positive Political Theory of Legislative History: New Perspectives on the 1964 Civil Rights Act and Its Interpretation*, in *University of Pennsylvania Law Review*, 151, 2003, p. 1417 ss.

RODWAY S., *Just How Fair Will Processing Notices Need to Be under the GDPR*, in *Privacy & Data Protection*, 16, 3, 2016

ROGERS C., *Client-Centered Therapy: ItsCurrentPractice, Implications and Theory*, Londra, 1951

ROLLA G., *La tutela dei diritti fondamentali*, Roma, 2012

ROLNICK D., DONTI P.L., KAACK L.H., KOCHANSKI K., LACOSTE A., SANKARAN K. ET AL., *TacklingClimateChange with Machine Learning*, in *ACM Computing Surveys*, 55, 2, 42, 2019

ROMM T., HARWELL D., STANLEY-BECKER I., *Facebook bans deepfakes, but new policy could not cover controversial Pelosi video*, The Washington Post, 7 gennaio 2020.

RONGA G., *Le varie ipotesi di responsabilità cosiddetta da "contatto sociale"*, in VIOLA L. (A CURA DI), *La responsabilità civile ed il danno*, Milano, 2007, 1, p. 90 ss.

ROSEN J., *The unvanted gaze. The destruction of privacy in America*, New York, 2000

ROSEN J., *The right to be forgotten*, in *Stanford Law Review Online*, 88, 64, 2012

ROSENBERG E. *Facebook blocked many gay-themed ads as part of its new advertising policy, angering LGBT groups*, The Washington Post, 3 ottobre 2018

ROSENBLATT F., *The Perceptron - a perceiving and recognizing automaton*, Report 85-460-1 - Cornell Aeronautical Laboratory, Buffalo (US), 1957

ROSENFELD A., ZEMEL R., TSOTSOS J.K., *The Elephant in the Room*, 2018, https://arxiv.org/abs/1808.03305

ROSENFELD M., *Affirmative action, justice, and equalities: a philosophical and constitutional appraisal*, in *Ohio State Law Review*, 46, 1985, p. 845 ss.

ROSENFELD M., SAJO A., *Spreading liberal constitutionalism: an inquiryinto the fate of free speech rights in new democracies*, in CHOUDRY S. (A CURA DI), *The migration of constitutionalideas*, Cambridge, 2007

ROSER M., RITCHIE H., *Technological progress*, in *Our World in Data*, 2020, https://ourworldindata.org/technological-progress

ROSIE G., *Google and advertising: digital capitalism in the context of post-fordism, the reification of language, and the rise of fake news*, in *Palgrave communications*, 3, 1,2018 p. 1 ss.

ROSSI S., *Contatto sociale (fonte di obbligazione)*, in *Digesto delle discipline privatistiche*, sez. civile, Appendice di aggiornamento V, Torino, 2010, p. 346 ss.

RUBIN V., CHEN Y., CONROY N., *Deception detection for news, three kinds of fakes*, in *Proceedings of the American Society for Information Science and Technology*, 52, 2015

RUDIN C., *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead*, in *Nature Machine Intelligence*, 1, 5, 2019, p. 206 ss.

RUFFINI GANDOLFI M.L., *Il diritto all'identità personale di fronte alla Corte Suprema degli Stati Uniti (il tort di false light in the public eye)*, in *Riv. dir. ind.*, 1981, p. 237 ss.

RUFFOLO U., *L'Intelligenza artificiale in sanità: dispositivi medici, responsabilità e "potenziamento",* in *Giurisprudenza italiana*, 2,2021, p. 502 ss.

RULE C., *Online Dispute Resolution and the Future of Justice*, in *Annual Review of Law and Social Science*, 16, 1, 2020

RUMELHART D.E., HINTON G.E., WILLIAMS R.J., *Learning representations by back-propagating errors*, in *Nature*, 323, 6088, 1986, p. 533 ss.

RUMELHART D.E., MCCLELLAND J.L. (a cura di), *Paralleldistributed processing*, Cambridge (US), 1986

RUSSELL S., NORVIG P., *Artificial intelligence: a modern approach (4ª ed.)*, Hoboken (NJ), 2021

RUTKIN A.H., *The tiny changes that can cause AI to fail*, BBC – Future Now, 11 aprile 2017, https://www.bbc.com/future/article/20170410-how-to-fool-artificial-intelligence (16 maggio 2022)

SALAMANCA R., *La gripecolapsaloshospitales de media España*, El Mundo, 12 gennaio 2017

SALVADORI L., VILLI C., *Il luddismo. L'enigma di una rivolta*, Sesto S. Giovanni, 1987

SAMEK W., MÜLLER K.R., *Towards Explainable Artificial Intelligence*, in SAMEK W., MONTAVON G., VEDALDI A., HANSEN L., MÜLLER K.R. (A CURA DI), *Explainable AI: interpreting, explaining and visualizing deep learning*, Cham, 2019

SAMPLE I., *My poker face: AI wins multiplayer game for first time*, The Guardian, 11 luglio 2019

SAMUEL A., *Some studies in machine learning using the game of checkers*, in *IBM Journal*, 3, 1959, p. 211 ss.

SAN-HUN C., *Google's computer program beats Le-Sedol in Go tournament*, The New York Times, 16 marzo 2016

SANNA I., *Diritto di cittadinanza e uguaglianza sostanziale*, Roma, 2014

SANNA P., *Il regime di responsabilità dei providers intermediari di servizi della società di informazione*, in *Responsabilità civile e previdenza*, 1, 2004, p. 279 ss.

SANTO V., *La società moderna e le nuove esigenze di accesso alla giustizia*, in *Studia Prawnoustrojove*, 24, 2014, p. 269 ss.

SANTORO PASSARELLI F., *Dottrine generali del diritto civile*, Napoli, 1966, p. 50 ss.

SANTOSUOSSO A., *The human rights of nonhuman artificial entities: an oxymoron?*, in *Jahrbuch für Wissenschaft und Ethik*, 19, 2015, pp. 203-237

SANTOSUOSSO A., *Intelligenza Artificiale e Diritto : Perché Le Tecnologie Di IA Sono Una Grande Opportunità per Il Diritto*, Milano, 2020

SANTOSUOSSO A., *Intelligenza artificiale, conoscenze neuroscientifiche e decisioni giuridiche*, in *Teoria e critica della regolazione sociale*, 1 2021, p. 189 ss.

SARAT A., *". . . The lawisall over": power, resistance and the legalconsciousness of the welfare poor*, in EWICK P. (A CURA DI),*Consciousness and deology*, Londra, 2006 p. 37 ss.

SARKAR S. ET AL., *Accuracy and Interpretability Trade-offs in Machine Learning Applied to Safer Gambling*, in *CEUR Worskshop Proceedings*, 2016, p. 1773 ss.

SARTOR G., BRANTING K. (A CURA DI), *Judicial Applications of Artificial Intelligence*, Dordrecht, 1998

SARTOR G., *Search engines as controllers. Inconvenientimplications of a questionableclassification*, in *Maastricht Journal of European Comparative Law*, 21, 3, 2014, p. 564 ss.

SASSÓLI M., *Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified*, in *International law studies*, 90, 2014, p. 308 ss.

SATARIANO A., ISAAC M., *The silent partner cleaning up Facebook for $500 million a year*, The New York Times, 31 agosto 2021

SCACCIA G., *Proporzionalità e bilanciamento tra diritti nella giurisprudenza delle corti europee*, in *Rivista AIC*, 3, 2017

SCAFFARDI L., *La medicina alla prova dell'intelligenza artificiale*, in *DPCE – Online*, 1, 2022, p. 349 ss.

SCHAFFER J., *Solving the Game of Checkers*, in *Mathematical Sciences Research Institute Publications*, 29, 1996, p. 119 ss.

SCHARFENBERG D., *Why Facebook and Google shouldpayyou for your data*, The Boston globe, 14 giugno 2018

SCHERMER B.W., CUSTERS B., VON DER HOF S., *The crisis of consent: how stronger legal protection may lead to weaker consent in data protection*, in *Ethics and information technology*, 16, 2014, p. 171 ss.

SCHEMMER M., HEMMER P., KÜHL N., BENZ C., SATZGER G., *Should I Follow AI-based Advice? Measuring Appropriate Reliance in Human-AI Decision-Making*, in *Conference on Human Factors in Computing Systems 2022*, *Workshop on trust and reliance in AI-human teams (trAIt)*, New Orleans, 2022

SCHMITZ A.J., *Expanding access to remediesthrough e-court initiatives*, in *Buffalo Law Review*, 67, 1, 2019, p. 89 ss.

SCHRETTENBRUNNER M.B., *Artificial-Intelligence-Driven Management*, in *IEEE Engineering Management Review*, 48, 2, 2020

SCHUCHMANN S., *History of the first AI winter*, in *Towards data science*, 2019

SCHWAB K., *The fourth industrial devolution*, New York, 2017

SCHWARTZ P., *Risk and high-risk: Walking the GDPR tighrope*, in *IAPP*, 29 marzo 2016

SCIACCA F., *I diritti sociali alla prova*, in *Rivista di filosofia del diritto*, 1, 2022, p. 182 ss

SCURICH N., *The case against categorical risk estimates*, in *Behavioral Sciences & the Law*, 36, 5, 2018

SEARLE J.R., *Minds, brains and programs*, in *The behavioral and brain sciences*, 3, 1980, p. 417 ss.

SEARLE J.R., *Is the brain's mind a computer program?,* in *Scientific American*, 262, 1, 1990, p. 26 ss.

SELBST A.D., POWLES J., *Meaningful information and the right to explanation*, in *International data privacy law*, 7, 4, 2017

SELMI M., *Algorithms, discrimination and the law*, *Ohio State Law Journal,* 82, 4, p. 611 ss.

SERBAN A., SANKAR C., GERMAIN M., ZHANG S., LIN Z., SUBRAMANIAN S. ET AL., *A Deep Reinforcement Learning Chatbot*, 2017, http://arxiv.org/abs/1709.02349

SERPICO D., *Esiste davvero l'intelligenza generale? Prospettive delle scienze cognitive*, in *NeaScience*, 9, 2, 2015, p. 216 ss.

SHAIKH S.J., *Artificial Intelligence and resourceallocation in health care: The process-outcome divide in perspectives on moral decision-making*, in *Ceur – WP*, 2884, 112, http://ceur-ws.org/Vol-2884/paper_122.pdf

SHAPIRO S.H., *The fifth generation project – a trip report*, in *Communications of the ACM*, 26, 9, 1983, p. 637 ss.

SHASHIKUMAR S.P. ET AL., *Development and Prospective Validation of a Deep Learning Algorithm for Predicting Need for Mechanical Ventilation*, in *Chest*, 159, 6, 2021

SHAW J.A., SETHI N., BLOCK B.L., *Five things every clinician should know about AI ethics in intensive care*, in *Intensive Care Medicine*, 47, 2, 2021, p. 157 ss.

SHEN T., LIU R., BAI J., LI Z., *"deep fakes" using generative adversarial networks (gan)*, 2018

SHERMAN E., *Privacy policies are great – For PhDs*, in *CBS News*, 4 Sept. 2018

SHERREL L., *Combinatorial explosion*, in RUNEHOV A.L.C., OVIEDO L. (a cura di), *Encyclopedia of Sciences and Religions*, Dordrecht, 2013, doi:10.1007/978-1-4020-8265-8_201037

SHIEBER J., *After criticism over moderation treatment, Facebook raises wages and boosts support for contractors*, TechCrunch+, 13 maggio 2019

SHNEIDERMAN B., *Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy*, in *International Journal of Human–Computer Interaction*, 36, 6, 2020

SHOKRI R.,STROBEL M.,ZICK Y., *On the Privacy Risks of Model Explanations*, 2019, https://arxiv.org/abs/1907.00164

SHORTLIFFE E.H., BUCHANAN B.G., *A model of inexact reasoning in medicine*, in *Mathematical Biosciences*, 23, 3-4, 1975, p. 351-379

SHRESTHA Y.R., BEN-MENAHEM S.M., VON KROGH G., *Organizational Decision-Making Structures in the Age of Artificial Intelligence*, in *California Management Review*, 61, 4, 2019, p. 66–83

SIAARTI, *Raccomandazioni di etica clinica per l'ammissione a trattamenti intensivi e per la loro sospensione in condizioni eccezionali di squilibrio tra necessità e risorse disponibili*, 6 marzo 2020

SIAARTI-SMILA, *Decisioni per le cure intensive in caso di sproporzione tra necessità assistenziali e risorse disponibile in corso di pandemia di Covid-19*, 13 gennaio 2021

SICCARDI C., *La "loi Avia". La leggefrancesecontrol'odio online (o quelloche ne rimane)*, in D'AMICO M., SICCARDI C., *La Costituzione non odia: conoscere, prevenire, contrastarel'hate speech online*, Torino, 2021

SILBERG J., MANYIKA J., *Notes from the AI frontier: Tackling bias in AI (and in humans)*, in *McKinsey Global Institute*, 2019

SILVA S., KENNEY M., *Algorithms, Platforms, and Ethnic Bias*, in *Communications of the ACM*, 62, 11,p. 37 ss.

SILVA S., KENNEY M., *Algorithms, Platforms, and Ethnic Bias: An Integrative Essay*, in *Phylon*, 55, 1-2, 2018, p. 9–37

SIMKOFF M., MAHDAVI A., *AI doesn't actually exist yet*, in *Scientific American – Observations*, 12 novembre 2019

SIMONCINI A., *L'algoritmo incostituzionale: l'intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019, p. 63 ss.

SIMONCINI A., *Profili costituzionali della amministrazione algoritmica*, in *Rivista trimestrale di diritto pubblico*, 4, 2019

SIMONCINI A., *Amministrazione digitale algoritmica. Il quadro costituzionale*, in GALETTA D.U., CAVALLO PERIN R. (A CURA DI), *Il diritto dell'Amministrazione Pubblica digitale,* Torino, 2020, p. 1-41

SINAGRA E., ROSSI F., RAIMONDO D., *Use of Artificial Intelligence in Endoscopic Training: Is Deskilling a Real Fear?*, in *Gastroenterology* 160, 6, 2021, p. 2212 ss.

SINGER M. ET AL., *The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)*, in *Journal of American Medical Association*, 315, 8, 2016, p. 801-810

SINGH S., *Everything in Moderation. An Analysis of How Internet Platforms Are Using Artificial Intelligence to Moderate User-Generated Content*, New America – Open Technology Institute, 2019

SINGHAL A., SINHA P., PANT R., *Use of Deep Learning in Modern Recommendation System: A Summary of Recent Works*, in *International Journal of Computer Applications*, 180, 7, 2017, p. 17-22

SLAGE J.R., *A heuristic program that solves symbolic integration problems in freshman calculus: symbolic automatic integrator (SAINT)*, 1961, https://dspace.mit.edu/handle/1721.1/11997

SŁOWIK A., BOTTOU L., *Algorithmic Bias and Data Bias: Understanding the Relation between Distributionally Robust Optimization and Data Curation*, 2021, https://arxiv.org/abs/2106.09467

SMILANSKY S., *Free Will and Illusion*, Oxford, 2000

SMITH M. D., VAN ALSTYNE M., *It's time to update Section 230*, in *Harvard Business Review*, 12 agosto 2021

SMITH V., *Rationality in economics: constructivist and ecological forms*, Leiden, 2007

SMUHA N.A. ET AL., *How the EU Can AchieveLegallyTrustworthy AI: A Response to the EuropeanCommission'sProposal for an Artificial Intelligence Act*, 2021

SNJIDERS C., MATZAT U., REIPS U.D., *"Big Data": Big Gaps of Knowledge in the Field of Internet Science*, in *International Journal of Internet Science*, 7, 2012, p, 1-5

SOLOVE D.J., *Privacy self-management and the consent dilemma*, in *Harvard Law Review*, 126, 2013, p. 1880 ss.

SOLOVE D., SCHWARTZ P.M., *Information privacy law (7th ed.)*, New York, 2021

SOMMA A., *I diritti della personalità e il diritto generale della personalità nell'ordinamento privatistico della Repubblica Federale Tedesca*, in *Rivista trimestrale di diritto e procedura civile*, 3, 1996, p. 805 ss.

SORRENTINO F., *Eguaglianza formale*, in *Costituzionalismo.it*, 3, 2017

SORRENTINO F., *Eguaglianza*, Torino, 2011

SOUCRAMANIEN F.M., *Le principe d''égalité dans la jurisprudenceduConseilconstitutionnel*, Marsiglia, 1999

SOURDIN T., CORNES R., *Do JudgesNeed to Be Human? The Implications of Technology for Responsive Judging*, inSOURDIN T., ZARISKI A. (A CURA DI), *The Responsive Judge: International Perspectives*, Singapore, 2018, p. 87 ss.

SOURDIN T., *Judge v Robot?:Artificial intelligence and judicialdecision-making*, in *The University of New South Wales Law Journal*, 4, 41, 2018, p. 1114 ss.

SPATARO O., *Stato di emergenza e legalità cosituzionale alla prova della pandemia*, in *Federalismi.it*, 11, 2022, p. 158 ss.

SPEITH T., *A review of taxonomies of expleinable artificial intelligence (XAI) methods*, in *2022 ACM Conference on fairness, accountability, and transparency*, p. 2239 ss.

SPINSANTI S., *Il Covid-19 ci ha costretto a confrontarci col triage*, Corriere della Sera, 21 marzo 2022

STAFF T., *Facebook bans far-right political figures Marzel, Ben Ari from its platforms*, The Times of Israel, 11 agosto 2021

STAFF T., *Polish far-right party banned from Facebook over alledged COVID disinformation*, Euronews, 6 gennaio 2022

STALLA-BOURDILLON S., *Liability exemptions wanted! Internet intermediaries' liability under Uk law*, in *Journal of International Commercial Law and Technology*, 7, 4, 2012

STARK R., *Discovering God: the origins of the great religions and the evolution of belief*, New York, 2007, p. 10 ss.

STATT N., *Deep Mind's Star Craft 2 AI is now better than 99.8 percent of all human players*, The Verge, 30 ottobre 2019

STATT N., *Open AI's Dota 2 AI steamrolls world champion e-sports team with back-to-back victories*, The Verge, 13 aprile 2019

STEFFEK F., UNBERATH H., GENN H., GREGER R., MENKEL-MEADOW C., *Regulating Dispute Resolution: ADR and Access to Justice at the Crossroads*, Oxford, 2013

STEFFEL J., WILLIAMS E.F., *Delegating decisions: recruiting others to make choices we might regret*, in *Journal of Consumer Research*, 44, 5, 2018, p. 1015 ss.

STEIGER D., *Protecting Democratic Elections Against Online Influence via "Fake News" and Hate Speech – The French Loi Avia and Loi No. 2018–1202, the German Network Enforcement Act and the EU's Digital Services Act in Light of the Right to Freedom of Expression*, in KOTZUR

M., SCHIEDERMAIR S., STEIGER D., WENDEL M. (A CURA DI), *Theory and practice of the European Convention on Human Rights*, 2021

STELLA F., *Leggi scientifiche e spiegazione causale nel diritto penale*, Milano, 1990

STERNBERG R.J., *Intelligence*, in FREEDHEIM D.K., WEINER I.B. (a cura di), *Handbook of psychology: History of psychology*, Hoboken, 2013, p. 155 ss.

STONE G.R., BOLLINGER L.C., *The free speech century*, New York, 2019

STRACHEY C.S., *Logical or non-mathematical programs*, in *ACM '52: Proceedings of the 1952 ACM national meeting*, 1952, p. 46-49

STRADA G., *Pappagalli verdi. Cronache di un chirurgo di guerra*, Milano, 1999

STRADELLA E., *La regolazione della robotica e dell'intelligenza artificiale: il dibattito, le proposte, le prospettive. Alcuni spunti di riflessione*, in *Media Laws*, 1, 2019, p. 73 ss.

STRADELLA E., *Cancellazione e oblio: come la rimozione del passato, in bilico tra tutela dell'identità personale e protezione dei dati, si impone anche nella Rete, quali anticorpi si possono sviluppare, e, infine, cui prodest?*, in *Rivista AIC*, 4, 2016

STRICKE B., *People v. Robots: A Roadmap for Enforcing California's New Online Bot Disclosure Act*, in *Vanderbilt Journal of Entertainment and Technology Law, 4, 4, 6, 2020, p.* 839 ss.

STRINGHAM E.P. (a cura di), *Private governance: creating order in economic and social life*, Oxford, 2015

STRONG F.R., *Fifty Years of 'Clear and Present Danger': From Schenck to Brandenburg - and Beyond*, in *The Supreme Court Review*, 1969, p. 41 ss.

STUDDERT D.M. ET AL., *Defensive Medicine Among High-Risk Specialist Physicians in a Volatile Malpractice Environment*, in *JAMA*, 293, 21, 2005, p. 2609 ss.

SUBRAMANIAN N., ELHARROUSS O., AL-MAADEED S., CHOWDHURY M., *A review of deep learning-based detection methods for COVID-19*, in *Computers in Biology and Medicine*, 143, 2022, p. 105233 ss.

SUN Q., *China's social credit system was due by 2020 but is far from ready*, in *Algorithm Watch*, 2021, https://bit.ly/3pQavsC

SUNSTEIN C., *Democracy and the Problem of Free Speech*, New York, 1995

SUNSTEIN C., *Republic.com*, Princeton, 2001

SUNSTEIN C., *Republic.com 2.0*, Princeton, 2007

SUNSTEIN C., THALER R., *Nudge: improving decisions about health, wealth and happiness*, New Haven, 2008

SUNSTEIN C., *Nudging: a very short guide*, in *Journal of consumer policy*, 37, 2014, p. 583 ss.

SUNSTEIN C., *#Republic.com: divided democracy in the age of social media*, Princeton, 2017

SUSSER D., ROESSLER B., NISSENBAUM H., *Technology, autonomy and manipulation*, in *Internet policy review*, 8, 2, 2019, doi: 10.14763/2019.2.1410

SWEENEY L., *Onlyyou, your doctor and manyothersmay know*, in *Technology Science*, September 28, 2015

SWEENEY L., VON LOEWENFELDT M., PERRY M., *Saying it's anonymous doesn't make it so: re-identification of "anonymized" law school data*, in *Technology Science*, November 12, 2018

SWISS COGNITIVE – THE GLOBAL AI HUB, *Deep learning won't detect fake news, but it will give fact-checkers a boost*, 29 febbraio 2020, https://bit.ly/2yu7Q0k

TALCIN O., *5 significantreasonswhyexplainableAI i san existentialneeds for humanity*, in *Towards Data Science*, 28 dicembre 2020

TAMBIAH S.J., *Magic, science, religion, and the scope of rationality*, Cambridge, 1990

TAMIBINI D., *How advertising fuels fake news*, in *LSE Media policy blog*, 2017, https://bit.ly/3mT3uoM

TAN S., CARUANA R., HOOKER G., LOU Y., *Distill-and-Compare: Auditing Black-Box Models Using Transparent Model Distillation*, in A.A.V.V., *Proceedings of the Conference on AI, Ethics, and Society*, New Orleans, 2018, p. 303 ss.

TARUFFO M., *La motivazione della sentenza civile*, Padova, 1975

TAYLOR C.R., KITCHEN P.J., SARKEES M.E., LOLK C.O., *Addressing the Janus face of customer service: a typology of new age service failures*, in *European journal of marketing*, 54, 10, 2020

TEGA D., *I diritti in crisi. Tra Corti nazionali e Corte europea di Strasburgo*, Milano, 2012

TEMPERTON J., *How the 5G coronavirus conspiracy theory torethrough the internet*, Wired, 6 aprile 2020

THALER R.H., SUNSTEIN C.R., *Nudge. The final edition*, New York, 2021, p. 91 ss.

The new rules of the "creator economy", The Economist, 8 maggio 2021

The rise of the influencer economy, The Economist, 2 aprile 2022

The top players in the AI-powered contract management space, www.cenza.co, 26 maggio 2022

The toughethicaldecisions doctors face with covid-19, The Economist, 2 aprile 2020

THOMAS A.K., *The history of radiology*, Oxford, 2013

THOMAS D.B., *The Alveyprogramme – intelligent knowledge-based systems aspects*, in *R&D Management*, 15, 2, 1985, p. 101-103

THORTON S., *Delivering Faster Results with Food Inspection Forecasting*, in *Data-smart city solutions*, 19 maggio 2015, https://bit.ly/3qZluA4 (20 agosto 2022)

TIMAN T., MANN Z., *Data protection in the era of artificial intelligence: trends, existing solutions and recommendations for privacy-preserving technologies*, in CURRY E. ET AL., *The elements of big data value. Foundations of the research and innovation ecosystem*, Cham, 2021, p. 153-177.

TOPOL E., *Deep medicine. How artificial intelligence can make healthcare human again*, New York, 2019

TOULMIN S., *The uses of argument*, Cambridge, 1958

TRAVERSO P., *Breve introduzione tecnica all'intelligenza artificiale*, in DPCE Online, 51, 1, 2022, p. 155-167

TREMLET G., *How didSpaingetits Coronavirus response so wrong?*, The Guardian, 26 marzo 2020

TREMOLADA L., *Cosa sono i data trust e perchè possono aiutare la privacy e la società civile*, il Sole 24 Ore, 30 luglio 2021

TRUCCO L., *Introduzione allo studio dell'identità individuale nell'ordinamento costituzionale italiano*, Torino, 2004

TRUOG R.D., MITCHELL C., DALEY G.Q., *The Toughest Triage — AllocatingVentilators in a Pandemic*, in *The New England Journal of Medicine*, 382, 2020, p. 1973 ss.

TSFATI Y., BOOMGAARDEN H.G., STRÖMBÄCK J., VLIEGENTHART R., DAMSTRA A., LINDGREN E., *Causes and consequences of mainstream media dissemination of fake news: literature review and synthesis*, in *Annals of the International Communication Association*, 44, 2, 2020, p. 157 ss.

TUCKER C., *Privacy, algorithms and artificial intelligence*, in *The economics of artificial intelligence: an agenda*, Chicago, 2019, p. 423 ss.

TURING A., *Computing machinery and intelligence*, in *Mind*, 236, 1950, p. 433 ss.

TUZET G., *Dover decidere. Diritto, incertezza, ragionamento*, Roma, 2010

TWITTER, *Permanent suspension of @realDonaldTrump*, 8 gennaio 2021

U.S. Capitol riot, New York Times (online), https://nyti.ms/2TL1iEY

U.S. FOOD AND DRUG ADMINISTRATION, HEALTH CANADA, UK'sMEDICINES AND HEALTHCARE PRODUCTS REGULATORY AGENCY, *Good Machine Learning Practice for Medical Device Development: GuidingPrinciples*, 2021

UCKELMANN D., HARRISON M., MICHAHELLES F. (a cura di),*Architecting the Internet of Things*, Berlin-Heidelberg, 2011.

UK CIVIL JUSTICE COUNCIL, *Online Dispute Resolution for low value civil claims*, 2015

UK GOVERNMENT – DEPARTMENT FOR DIGITAL, CULTURE, MEDIA AND SPORT, *AI sector deal*, 26 aprile 2018

UK GOVERNMENT, *Uk pandemic preparedness*, Policy Paper, 5 novembre 2020

UK GOVERNMENT – DEPARTMENT OF HEALTH AND SOCIAL CARE, *£36 million boost for AI technologies to revolutionise NHS care*, 16 giugno 2021

UK GOVERNMENT, *National artificial intelligence (AI) strategy*, 22 settembre 2021

UK GOVERNMENT, *National AI Strategy – AI Action Plan*, diffuso il 18 luglio 2022

UNESCO AD HOC EXPERT GROUP (AHEG), *First draft of the recommendation on the ethics of artificial intelligence cit.*, Parigi, 7 settembre 2020, p. 4

UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, 24 novembre 2021, 41 C/73 (Annex)

USERTESTING, *healthcare chatbot apps are on the rise but the overall custode experience (cx) falls short according to a UserTesting report*, 2019, https://bit.ly/3ATovGM

VALASTRO A., *Principi comuni a livello europeo in materia di propaganda elettorale televisiva*, in *Quaderni costituzionali*, 1, 1997, p. 109-130

VALDES E., *Biolaw, geneticharm and fourth generation human rights*, in *Boletínmexicano de derechocomparado*, 48, 144, 2015, p. 1197-1228

VALLE-CRUZ D., ALEJANDRO RUVALCABA-GOMEZ E., DE SOUSA W., DE MELO E.R.P., BERMEJO P.H.D.S., FARIAS R., GOMES A.O., *How and whereisartificial intelligence in the public sectorgoing? A literature review and research agenda*, in *Government Information Quarterly*, 36, 4, 2019

VALLE-CRUZ D., RUVALCABA-GOMEZ E.A., SANDOVAL-ALMAZAN R., IGNACIO CRIADO A., *A Review of Artificial Intelligence in Government and its Potential from a Public Policy Perspective*, in *Proceedings of the 20th Annual International Conference on Digital Government Research*, Dubai United Arab Emirates, 2019, p. 91-99

VALLOR S., *Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character*, in *Philosophy & Technology*, 28, 1, 2015, p. 107-124

VALLOR S., *The future of military virtue: autonomous systems and the moral deskilling of the military*, in *2013 5th International Conference on Cyber Conflict (CYCON 2013)*, 2013, https://bit.ly/3qU4URV

VAN DER DONK B.*,The freedom to conduct a business as a counterargument to limit platform users' freedom of expression*, in HINDELANG S., MOBERG A., *YSEC Yearbook of Socio-Economic Constitutions 2021: Triangulating Freedom of Speech*, Berlino, 2022, p. 33–58

VAN DER VOORT H.G., KLIEVINK A.J., ARNABOLDI M., MEIJER A.J., *Rationality and politics of algorithms. Will the promise of big data survive the dynamics of public decision making*, in *Government Information Quarterly*, 36, 1, 2019

VAN EECKE P., *Online Service Providers and Liability: A Plea for a BalancedApproach*, in *Common Market Law Review*, 48,2011,p. 1455 ss.

VARI F., *L'affermazione del principio di eguaglianza nei rapporti tra privati. Profili costituzionali*, Torino, 2017

VARIAN H.R., *Beyond big data*, NABE Annual Meeting – San Francisco, 10 settembre 2013

VASAK K., *A 30 years struggle. The sustained efforts to give force of law to the Universal Declaration of Human Rights*, The UNESCO Courier, Nov. 1977, p. 29 ss.

VENKATESH A. ET AL., *On Evaluating and Comparing Open Domain Dialog*, 2018, http://arxiv.org/abs/1801.03625

VERBA S., ORREN G.R., *The meaning of equality in America*, in *Political Science Quarterly*, 100, 3, 1985, p. 369-387

VIGLIETTA G., *Il giudice penale e i diritti fondamentali*, in *Questione giustizia*, 5, 2007

VIGORITO A., *Piattaforme digitali e "political speech": dal caso Facebook-CasaPound alla vicenda Trump-Twitter*, in *giustiziacivile.com*, 11, 2020, p. 16 ss.

VIJAYAN V., CONNOLLY J.P., CONDELL J., MCKELVEY N., GARDINER P., *Review of Wearable Devices and Data Collection Considerations for Connected Health*, in *Sensors*, 21, 16, 2021, p. 5589 ss.

VILLASCHI P., *Facebook come la RAI? Note a margine dell'ordinanza del Tribunale di Roma del 12.12.2019 sul caso Casa Pound c. Facebook*, in *Osservatorio AIC*, 2, 2020, p. 430 ss.

VILONE G., LONGO E., *Explainable artificial intelligence: a systematic review*, 2020, arXiv:2006.00093

VILONE G., LONGO E., *Notions of explainability and evaluation approaches for explainable artificial intelligence*, in *Information Fusion*, 76, 2021, p. 89–106

VILONE G.; LONGO L., *Classification of Explainable Artificial Intelligence Methods through Their Output Formats*, in *Machine Learning and Knowledge Extraction*, 3, 3, 2021, p. 615-661.

VINCENT J.L. ET AL., *The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure*, in *Intensive Care Medicine*, 22, 7, 1996

VINGE V., *The Coming Technological Singularity: How to Survive in the Post-Human Era*, in *Vision-21 Interdisciplinary Science and Engineering in the Era of Cyberspace - Proceedings of a symposium cosponsored by the NASA Lewis Research Center and the Ohio Aerospace Institute and held in Westlake (Ohio)*, 1993, p. 11-22

VIOLANTE L., *Diritto e potere nell'era digitale. Cybersociety, cybercommunity, cyberstate, cyberspace: tredici tesi*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2022, p. 145-153

VIRGA P., *Libertà giuridica e diritti fondamentali*, Milano, 1947

VOLPE G., *Il costituzionalismo del Novecento*, Roma-Bari, 2000

VON GAREIS K., *DerAllgemeineTeildesBurgerlichenGesetzbuchs*, Berlino, 1900

VON GIERKE O., *AllgemeinerTeil und Personenrecht*, Lipsia, 1895

VON OTTERLO M., *A machine learning view on profiling*, in *Privacy, due process and the computational turn: the philosophy of law meets the phylosophy of technology*, New York, 2013, p. 41-65

VON ROOKHUIJZEN M., DE VET E., *Nudging healthy eating in Dutch sports canteens: a multi-method case study*, in *Public Health Nutrition*, 2020, doi:10.1017/S1368980020002013

VRENDENBURGH K., *The right to explanation*, in *The Journal of Political Philosophy*, 30, 2, 2022, p. 209 ss.

WACHTER S., MITTELSTADT B., FLORIDI L., *Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation*, in *International Data Privacy Law*, 7, 2, 2017, p. 76 ss.

WAHLSTEN D., *The Theory of Biological Intelligence: History and a Critical Appraisal*, in STERNBERG R.J., GRIGORENKO E.L., *The general factor of intelligence*, New York, 2002

WAKEFIELD J., *Christchurch shootings. Social media races to stop attack footage*, BBC News, 16 marzo 2019

WAKEFIELD J., *The hobbyistscompeting to make AI human*, in BBC news (online), 13 settembre 2019

WALLACH O., *How big tech makes theirbillions*, in *Visual capitalist*, https://www.visualcapitalist.com/how-big-tech-makes-their-billions-2020/ (1 marzo 2022)

WALLER A.D., *A Demonstration on Man of Electromotive Changes accompanying the Heart's Beat*, in *The Journal of Physiology*, 8, 5, 1887, p. 229-234

WALSH B., *Covid-19: The history of pandemics*, in *BBC future*, 26 marzo 2020, https://bbc.in/3VdS1QL (9 ottobre 2022)

WALTERS R., TRAKMAN L., ZELLER B., *Data protection law. A comparative analysis of Asia-Pacific and European approaches*, Singapore, 2019, p. 79-81

WANG F., CASALINO L., KHULLAR D., *Deep Learning in Medicine—Promise, Progress, and Challenges*, in *JAMA Internal Medicine*, 179, 3, 2019, p. 293 ss.

WANG F.F., *Online Dispute Resolution: Technology, management and legal practice from an international perspective*, Oxford, 2008

WANG P., *On defining artificial intelligence*, in *Journal of General Artificial Intelligence*, 10, 2, 2019, p. 1-37

Warning: Germany edges toward Chinese-style rating of citizens, Handelsblatt Global Edition, 17 February 2018.

WARREN S., BRANDEIS L., *The right to privacy*, in *Harvard Law Review*, 4, 5, 1890, p. 193-220

WARTMAN S.A., COMBS C.D., *Medicaleducation must move from the information age to the age of artificial intelligence*, in *Academic Medicine*, vol. 93, 8, 2018

WASHINGTON A.L., *How to Argue with an Algorithm: Lessons from the COMPAS-ProPublica Debate*, in *Colorado Technology Law Journal*, 17, 2018, p. 131 ss.

WEAVER J.B., *We Need the California Bot Bill, but We Need It to Be Better*, in *RAIL: The Journal of Robotics, Artificial Intelligence & Law*, 1, 6, 2018, p. 431 ss.

WEBER B., *Swift and slashing, computer topples Kasparov*, The New York Times, 12 maggio 1997

WEI T., CHEN X., LI X., ZHU Q., *Model-based and data-driven approaches for building automation and control*, in *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2019, doi:10.1145/3240765.3243485

WEINBERGER D., *Our machines now have knowledge we will never understand*, in *Wired*, 10 aprile 2017

WEISS S., KULIKOWSKI C., SAFIR A., *Glaucoma consultation by computer*, in *Computers in Biology and Medicine*, 8, 1, 1978, p. 25 ss.

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

WENG S., REPS J., KAI J., GARIBALDI J.M., QURESHI N,,*Can machine-learning improve cardiovascular risk prediction using routine clinical data?*, in *PLOS ONE*, 12, 4, 2017

WEST D.M., *What happens if robots take the jobs? The impact of emerging technologies on employment and public policy*, in *Center of Technology Innovation at Brookings*, Oct. 2015

WHITEHEAD P.B., HERBERTSON R.K., HAMRIC A.B., EPSTEIN E.G., FISHER J.M., *Moral distress among healthcare professionals: report of an institution-wide survey*, in *Journal of Nursing Scholarship*, 47, 2015, p. 117 ss.

WHITEN A. (A CURA DI), *Natural Theories of Mind: Evolution, Development, and Simulation of Everyday Mindreading*, Oxford, 1991

WHITTEN-WOODRINGETAL J., *Poison If You Don't Know How to Use It: Facebook, Democracy, and Human Rights in Myanmar*, in *The International Journal of Press/Politics*, 25, 3, 2020, p. 407 ss.

WILKERSON L., *Still waters run deep(fakes): the rising concerns of "deepfake" technology and its influence on democracy and the first amendment*, in *Missouri Law Review*, 86, 1, 2021, p. 407 ss.

WILKINSON J.H., *The Supreme Court, the EqualProtectionClause, and the Three Faces of ConstitutionalEquality*, in *Virginia Law Review*, 61, 5, 1975, p. 945 ss.

WILKS Y., STEVENSON M., *The grammar of sense: using part-of-speech tags as a first step in semantic disambiguation*, in *Natural Language Engineering*, 1998, 4, 2, p. 135 ss.

WILLIAMS Z., *Algorithms are taking over – and woe betide anyone they class as a 'deadbeat'*, The Guardian, 12 luglio 2018

WINFIELD A.T. ET AL., *Robot AccidentInvestigation: A Case Study in Responsible Robotics*, in CAVALCANTI A., DONGOL B., HIERONS R., TIMMIS J., WOODCOCK J. (A CURA DI), *Software Engineering for Robotics*, Cham, 2021, p. 165 ss.

WINIKOFF M., SARDELIC J., *Artificial Intelligence and the Right to Explanationas a Human Right*, in *IEEE Internet Computing*, 25, 2, 2021, p. 116-120

WISCHMEYER T., *Artificial Intelligence and Transparency: Opening the Black Box*, in WISCHMEYER T., RADEMACHER T. (A CURA DI), *Regulating Artificial Intelligence. Springer*, Cham, 2020 https://doi.org/10.1007/978-3-030-32361-5_4

WISEMAN J., GOLDSMITH S., *10 great ways data can make government better*, in *Data-smart city solutions*, 11 maggio 2017

WISEMAN J., GOLDSMITH S., *Optimizing the Quality and Delivery of City Emergency Medical Services | Data Science for Social Good Fellowship*, in *Data science for social good*, 2021

WOLTERS P.T.J., *The territorial effect of the right to be forgotten after Google v CNIL*, in *International Journal of Law and Information Technology*, 29, 1, p. 57-75

WOO M., *How Online Misinformation Spreads*, in *Knowable Magazine-Annual Reviews*, 2021

WOOLDRIDGE M., RAO A. (a cura di), *Foundation of rational agency*, Dordrecht, 1999

WOOLDRIDGE M., *Reasoning about rational agents*, Cambridge, 2003

WRIGLEY S., KLINEFELTER A., *Google LLC vs. CNIL: the location-based limits of the EU right to erasure and lessons for U.S. privacy law*, in *North Carolina Journal of Law & Technology*, 22, 4, 2021, p. 681 ss.

WU F., LU C., ZHU M., CHEN H., ZHU J., LI L., LI M., CHEN Q., LI X., CAO X., WANG Z., ZHA Z., ZHUANG Y., PAN Y., *Towards a new generation of artificial intelligence in China*, in *Nature Machine Intelligence*, 2, 2020, p. 312 ss.

WU J. ET AL., *Human-in-the-Loop Deep Reinforcement Learning with Application to Autonomous Driving*, arXiv, 2021, https://arxiv.org/abs/2104.07246

WU J., THORNE-LARGE J., ZHANG P., *Safety First: The risk of over-reliance on technology in navigation*, in *Journal of Transportation Safety & Security*, 2021, p. 1–28

WU X., XIAO L., SUN Y., ZHANG J., MA T., HE L., *A survey of human-in-the-loop for machine learning*, in *Future Generation Computer Systems*, 135, 2022, p. 364-381

XIAO-LIN L., ZHONG Z., *An overview of personal credit scoring: techniques and future work*, in *International Journal of Intelligence Science*, 2, 4A, 2012, doi: 10.4236/ijis.2012.224024

YABLONSKY S.A., *Multidimensional data-driven artificial intelligence innovation*, in *Technology innovation management review*, 9, 12, 2019, p. 16-28

YADAV D., SALMANI S., *Deepfake: A Survey on Facial Forgery Technique Using Generative Adversarial Network*, in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 2019, p. 852-857

YAMPOLSKIY R., *AI will fail, like everything else, eventually*, in *Mind matters news*, 14 luglio 2020

YAMPOLSKIY R., *Unexplainability and incomprensibility of artificial intelligence*, 2019, arXiv:1907.03869

YEUNG K., *"Hypernudge": Big Data as a mode of regulation by design*, in *Information, communication and technology*, 20, 1, 2017, p. 118 ss.

YEUNG K., *Why Worry about Decision-Making by Machine?*, in YEUNG K., LODGE M. (A CURA DI), *Algorithmic Regulation*, Oxford, 2019

YORK J.C., MCSHERRY C., *Content Moderation Is Broken. Let Us Count the Ways*, Electronic Frontier Foundation, 29 aprile 2019, https://bit.ly/2DEAafI

YU K., YANG Z., WU C., HUANG Y., XIE X., *In-hospital resource utilization prediction from electronic medical records with deep learning*, in *Knowledge-Based Systems*, 223, 2021, p. 107052 ss.

ZALNIERIUTE M., *Google LLC v. Commission nationale de l'informatique et des libertés (CNIL)*, in *American Journal of International Law*, 114, 2, 2020, p. 261 ss.

ZANDT F., *Big tech keepsgettingbigger*, in *Statista*, 29 ottobre 2019, https://www.statista.com/chart/21584/gafam-revenue-growth/

ZANELLATI P., *GDPR, l'approccio privacy basatosulrischio: come funziona e le misure da adottare*, in *Cybersecurity 360*, 4 dicembre 2020

ZANON N., *Fake news e diffusione dei social media: abbiamo bisogno di un'"Autorità Pubblica della Verità"?*, in *MediaLaws*, 1, 2018, p. 15 ss.

ZANZOTTO F.M., *Viewpoint: Human-in-the-loop Artificial Intelligence*, in *Journal of Artificial Intelligence Research*, 64, 2019, p. 243 ss.

ZAPPALÀ S., *La tutela internazionale dei diritti umani*, Bologna, 2011

ZARKADIS G., *"Data trusts" could be the key to better AI*, in *Harvard Business Review*, 10 novembre 2020

ZAVRŠNIK A., *Criminal justice, artificial intelligence systems, and human rights*, in *ERA Forum*, 20, 4, 2020

ZENO-ZENCOVICH V., *Onore*, *reputazione e identità personale*, in ALPA G., BESSONE M. (a cura di),*La responsabilità civile*, Torino, III, 1987 e *Identità personale*, in *Digesto delle discipline privatistiche*, IX, 1993, p. 294 ss.

ZENO-ZENCOVICH V., *Personalità (diritti della)*, in *Digesto delle discipline privatistiche*, XIII, 1995

ZHANG D., ZOU L., ZHOU X., HE F., *Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer*, in *IEEE Access*, 6, 2018, p. 28936 ss.

ZHANG J., *Application of machine learning in CT images and X-rays of COVID-19 pneumonia*, in *Medicine*, 100, 36, 2021

ZHOU L., GAO J., LI D., SHUM H., *The Design and Implementation of XiaoIce, an Empathetic Social Chatbot*, 2022, http://arxiv.org/abs/1812.08989

ZICCHITTU P., *I movimenti "antisistema" nell'agorà digitale: alcune tendenze recenti*, in *giurcost.org*, 5 marzo 2020

ZICCHITTU P., *La libertà di espressione dei partiti politici nello spazio pubblico digitale: alcuni spunti di attualità*, in *MediaLaws*, 2, 2021, p. 2 ss.

ZIMMERMAN D.L., *False Light Invasion of Privacy: The Light That Failed*, in *New York University Law Review*, 64, 1989, p. 364 ss.

ZUBOFF S., *The age of surveillance capitalism: the fight for a human future at the new frontier of power*, Londra, 2019

ZUIDERVEEN BORGESIUS F.J., *Discrimination, artificial intelligence, and algorithmic decision-making*, Council of Europe, Directorate General of Democracy, 2018

ZUIDERVEEN BORGESIUS F.J., *Strengthening legal protection against discrimination by algorithms and artificial intelligence*, in *The International Journal of Human Rights*, 24, 10, 2020, p. 1572-1593

41 minutes of fear: a video timeline from inside the Capitol siege, The Washington Post (online), https://wapo.st/3C2P3oA (10 maggio 2022)

80% of Russians Will Have State-Gathered 'Digital Profiles' by 2025, Official Says, in *The Moscow Times*, 28 settembre 2018

INDICE DELLA GIURISPRUDENZA CITATA

Italia

Stati Uniti d'America

Corte Europea dei Diritti dell'Uomo

Le sentenze della Corte EDU sono consultabili all'indirizzo:hudoc.echr.coe.int.