

## **L'AI Act e la trasparenza, tra tracciabilità, spiegabilità e conoscibilità. Un nuovo tassello nell'ecosistema regolatorio della ricerca?**

**Carlo Casonato, Marta Fasan, Marta Tomasi**

Università degli Studi di Trento

### SCOPO DELLO SCRITTO

Nell'aprile del 2021, l'Unione Europea e i suoi Stati Membri hanno iniziato a discutere su una prima versione di quello che, nel giugno del 2024, è stato approvato come il primo regolamento al mondo che, con approccio comprensivo ed orizzontale, stabilisce regole armonizzate sull'intelligenza artificiale (d'ora in poi AI Act<sup>1</sup>).

Questo contributo vuole sintetizzarne, in una prima parte, i passaggi più significativi, mettendone in luce alcuni punti di forza e di debolezza. In particolare, saranno descritte le ragioni alla base dell'approvazione dell'AI Act, alcune delle sue definizioni più importanti, gli ambiti compresi dal suo raggio d'azione, la sua logica e struttura di base. In una seconda parte, il contributo si rivolgerà al mondo della ricerca, tenendo presente che l'intelligenza artificiale (d'ora in poi IA) può esserne tanto l'oggetto di ricerca e sviluppo da parte di informatici, ingegneri informatici, roboticisti, ecc., quanto uno strumento utilizzabile da ricercatori di qualsiasi settore per raggiungere con maggior velocità o accuratezza i risultati delle proprie indagini. Che si dia il primo o il secondo caso, può essere interessante dare avvio a una riflessione intorno a un concetto cruciale tanto per la tecnologia in gioco, quanto per lo specifico ambito nel quale essa si trova collocata: la 'trasparenza'. Nonostante negli ultimi anni il dibattito scientifico abbia dato maggior risalto alla nozione di 'spiegabilità' quale possibile chiave di lettura da usare per affrontare i problemi di opacità tipici delle forme di IA più avanzate, il ricorso alla categoria giuridica

---

<sup>1</sup> Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio relativo all'intelligenza artificiale (IA) e che modifica alcune normative dell'Unione.

della ‘trasparenza’ offre, ad avviso di chi scrive, un più ampio ventaglio di soluzioni e strumenti per provare a rispondere alle numerose questioni poste dall’ingresso dell’IA nel settore in esame<sup>2</sup>. IA e ricerca, infatti, condividono una domanda comune rispetto a tale principio, il quale rappresenta, al contempo, una delle più complesse sfide per i nuovi sistemi intelligenti e uno degli irrinunciabili fondamenti dell’integrità della ricerca. Come si avrà modo di osservare, la ricerca che coinvolge l’IA rappresenta un campo di prova per la tenuta del principio di trasparenza che deve vestirsi di nuovi significati e declinazioni per restare fedele alla sua radice costitutiva.

### LE RAGIONI DELL’AI ACT

La base legale dell’AI Act è costituita (e non poteva essere altrimenti) dall’art. 114 del *Trattato sul Funzionamento dell’Unione Europea* che permette agli organi UE di adottare misure tese al riavvicinamento delle disposizioni degli Stati Membri «che hanno per oggetto l’instaurazione ed il funzionamento del mercato interno» (cons. 3)<sup>3</sup>. Da questo punto di vista, l’AI Act si inserisce all’interno del *New Legislative Framework* adottato dalla Commissione al fine di rendere più funzionale il mercato unico<sup>4</sup>, ponendosi come l’ennesimo tassello di una *product compliance regulation* tesa a scongiurare la frammentazione del mercato e favorire la certezza del diritto, in modo da sostenere la libera circolazione e l’innovazione. In tale contesto, rileva anche la natura dello strumento normativo prescelto, un regolamento il quale, a differenza di una direttiva, è direttamente applicabile in tutti gli Stati Membri.

---

<sup>2</sup> Come si avrà modo di osservare nelle prossime pagine, la scelta di adottare questa chiave di lettura trova giustificazione, da un lato, nella possibilità offerta dal principio di trasparenza di fornire soluzioni non solo in termini di spiegabilità, ma anche di tracciabilità e di conoscibilità dei sistemi di IA, con l’opportunità, quindi, di affrontare con più ampia portata i problemi che possono emergere dall’uso dell’IA nel contesto della ricerca. Dall’altro lato, è opportuno considerare come il riferimento al principio di trasparenza trovi spazio nei principali atti normativi elaborati e adottati in materia di IA, come dimostrato anche da quanto previsto dall’AI Act.

<sup>3</sup> In riferimento ai dati di carattere personale, i considerando menzionano anche la base legale rappresentata dall’art. 16 del TFUE, che ne dispone la protezione e libera circolazione.

<sup>4</sup> Tale approccio normativo si propone di migliorare la sorveglianza sul mercato per contrastare la commercializzazione di prodotti difettosi o che possano costituire un pericolo per la salute e l’ambiente. Prevede, inoltre, procedure per l’accreditamento di organi atti alla verifica di conformità dei prodotti oltre che le modalità di impiego del marchio CE. Cfr. European Commission, “New Legislative Framework”. [https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework\\_en](https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en).

Leggendo i numerosi ‘considerando’ (*recitals*) che anticipano l’articolato dell’AI Act (ben 180), emerge come il motivo per adottare una disciplina in materia non si limiti all’intenzione di assicurare regole uniformi per la circolazione di beni e servizi, ma derivi anche dal carattere del tutto peculiare dell’IA. La sua natura fortemente pervasiva e trasformativa, infatti, ne ha fatto e sempre più ne farà una componente centrale nello svolgimento di moltissime attività, con un impatto dirompente sugli individui e sulle nostre società, sui sistemi economici e di lavoro, sull’ambiente, sulle democrazie e sugli assetti di potere (pubblico e privato), sulla protezione dei diritti, oltre che sui possibili nuovi significati che potremo attribuire a termini come umanità e identità<sup>5</sup>. I motivi alla base dell’AI Act, quindi, non si riducono a garantire l’immissione nel mercato di prodotti sicuri, come capita per ogni tipo di prodotto certificato (con la marcatura CE), ma si spingono, nei limiti delle competenze dell’Unione, a indirizzarne lo sviluppo verso forme di IA affidabile e antropocentrica (*trustworthy and human-centered*), con il «fine ultimo di migliorare il benessere degli esseri umani» (cons. 6). La sfida dell’IA è di carattere propriamente esistenziale e l’AI Act tenta di affrontarla, per quanto possa farlo uno strumento giuridico europeo, in termini di protezione della salute, della sicurezza e dei diritti fondamentali, e puntando ad un rafforzamento della democrazia e dello Stato di diritto (cons. 8)<sup>6</sup>.

## DEFINIZIONI E RAGGIO D’AZIONE DELL’AI ACT

Uno dei primi compiti su cui gli esperti e i politici che hanno redatto l’AI Act hanno dovuto concentrare i propri sforzi è stato quello definitorio. Ne è scaturito un lungo articolo, il terzo, che contiene la descrizione di ben 68 voci.

Un sistema di IA, in primo luogo, è stato definito come «un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall’input che riceve come generare output quali previsioni,

---

<sup>5</sup> I riferimenti al riguardo sono innumerevoli. Fra gli altri, Luciano Floridi, *Etica dell’intelligenza artificiale. Sviluppì, opportunità, sfide*, Milano: Raffaello Cortina editore, 2022; Carlo Casonato, “Intelligenza artificiale e identità”, in *Enciclopedia Italiana Treccani*, Roma: Istituto Enciclopedia Italiana, XI appendice, vol. 2, 2024, 52-57.

<sup>6</sup> La tutela ambientale, rilevante in alcune versioni intermedie del progetto di AI Act e presente in alcuni considerando, non ha peraltro trovato un riconoscimento significativo nella versione definitiva dell’articolato.

contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali». In questo senso, allineandosi alla definizione già fornita dal Consiglio d'Europa, sono stati individuati quattro caratteri connotativi: un certo grado di (i) autonomia e di (ii) adattabilità, (iii) la capacità di influenzare il contesto e (iv) quella riconducibile alla deduzione (o inferenza). A questo riguardo, si sarebbe potuto fare maggior riferimento nella traduzione italiana dell'AI Act alla terminologia utilizzata nella versione inglese: in particolare, il termine *inference* avrebbe potuto essere più fedelmente tradotto utilizzando anche in italiano il verbo 'inferire'. Il considerando 12 dell'AI Act, su questa linea, riconduce proprio alle tecniche che consentono l'inferenza gli «approcci di apprendimento automatico che imparano dai dati come conseguire determinati obiettivi e approcci basati sulla logica e sulla conoscenza che traggono inferenze dalla conoscenza codificata o dalla rappresentazione simbolica del compito da risolvere. La capacità inferenziale di un sistema di IA trascende l'elaborazione di base dei dati consentendo l'apprendimento, il ragionamento o la modellizzazione».

Tale definizione pare di cruciale importanza, visto che consente di individuare quali saranno i sistemi che, soggetti al Regolamento, dovranno rispettarne i requisiti e quali invece, non presentando le quattro caratteristiche menzionate, ne saranno esenti.

Un secondo gruppo di definizioni, fra le molte che si potrebbero citare, riguarda le figure del fornitore e dell'utilizzatore. La prima riconduce al fornitore (*provider*) ogni persona fisica o giuridica, o qualsiasi organismo, che sviluppi o faccia sviluppare IA per finalità generali e la immetta nel mercato con il proprio nome o marchio, a titolo oneroso o gratuito<sup>7</sup>. Tale definizione è importante perché individua la figura cui è imposta la maggior parte degli obblighi previsti dall'AI Act, compresi quelli – vedremo – relativi alla costruzione di algoritmi che assicurino, nella misura del possibile, un certo grado di trasparenza. L'utilizzatore (*deployer*), invece, viene definito come la persona fisica o giuridica, o l'organismo, che utilizza un sistema di IA sotto la propria autorità. A questa figura sono imposti alcuni obblighi relativi, per quanto più rileva in questa sede, a una specifica declinazione della trasparenza art. 50, AI Act.

---

<sup>7</sup> La definizione riporta: «una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che sviluppa un sistema di IA o un modello di IA per finalità generali o che fa sviluppare un sistema di IA o un modello di IA per finalità generali e immette tale sistema o modello sul mercato o mette in servizio il sistema di IA con il proprio nome o marchio, a titolo oneroso o gratuito».

Fondamentale rilevanza assume anche l'individuazione del raggio d'azione, per così dire, dell'AI Act. A fronte della possibilità di stabilire regole diverse per i diversi ambiti di utilizzo, il Regolamento ha adottato un approccio comprensivo e orizzontale, destinato a disciplinare la creazione e l'utilizzo dei sistemi e modelli di IA in termini generalissimi.

Le regole previste, anzitutto, si applicano a tutta l'IA che voglia circolare sul territorio europeo. Ciò significa che i fornitori ne sono vincolati a prescindere dal fatto che siano stabiliti nell'Unione o in un paese terzo e che i loro prodotti debbano rispettare i requisiti richiesti dall'AI Act per il solo fatto di circolare all'interno del mercato europeo. In questo senso, l'Unione ha scommesso sul c.d. effetto Bruxelles: l'influenza esercitata in modo "discreto" a livello globale da parte della normativa europea riconducibile al fatto che tutti i produttori, anche di Paesi terzi, che non vogliono rinunciare a vendere all'interno del mercato unico dovranno fornire sistemi che siano rispettosi dei requisiti previsti dalla UE, adeguandosi così, in modo indiretto, anche ai relativi valori<sup>8</sup>.

Come anticipato, tutti i sistemi compresi nella definizione sopra illustrata sono oggetto del Regolamento, ad eccezione di quelli utilizzati in alcune aree e per alcuni scopi. L'AI Act, in primo luogo, non si applica, né potrebbe applicarsi, a settori che non rientrino nell'ambito di competenza del diritto dell'Unione (art. 2.3). Non essendo l'Unione Europea un'organizzazione dotata di competenza generale, a differenza degli Stati membri, il suo diritto si estende solo alle materie per cui i Trattati le conferiscono competenza, in conformità al principio di attribuzione (art. 5 TUE)<sup>9</sup>. Esclusi dagli obblighi previsti in generale per l'utilizzatore, sono, in secondo luogo, le persone fisiche che impiegano sistemi di IA nel corso di un'attività non professionale, di carattere puramente personale (art. 2.10). Fuori dal raggio d'azione dell'AI Act ricadono anche i sistemi utilizzati esclusivamente per scopi militari, di difesa o di sicurezza nazionale (art. 2.3)<sup>10</sup> e, di particolare importanza nella prospettiva di questo contributo e come si vedrà meglio più avanti, i sistemi

---

<sup>8</sup> Anu Bradford, *The Brussels effect: how the European Union rules the world*, Oxford: Oxford University Press, 2020.

<sup>9</sup> In questo senso, sono esclusi dall'AI Act i sistemi utilizzati, ad esempio, dalle pubbliche amministrazioni che svolgano attività su materie non coperte dal diritto dell'Unione o dai professionisti della salute all'interno della relazione medico-paziente.

<sup>10</sup> Maggiori dettagli nel considerando 24, AI Act.

e i modelli specificamente sviluppati e messi in servizio al solo scopo di ricerca e sviluppo scientifici (art. 2.6). Inoltre, non volendo correre il rischio di pregiudicare la libertà della scienza, il Regolamento esclude la propria applicabilità alle attività di ricerca, prova o sviluppo relative a sistemi di IA o modelli di IA prima della loro immissione sul mercato o messa in servizio. Tale eccezione, coerentemente con l'approccio basato sul rischio (su cui *infra*), non copre le prove in condizioni reali (art. 2.8)<sup>11</sup>.

#### LOGICA E STRUTTURA DELL'AI ACT

L'AI Act si muove all'interno di una logica basata sul rischio, con regole proporzionate rispetto al grado di pericolo emergente. Ciò significa che i sistemi sono divisi non sulla base delle loro caratteristiche tecniche (*Model-Based* v. *Machine Learning*, ad esempio) né prioritariamente sulla base degli ambiti di utilizzo, ma – appunto – sulla base della gravità di rischio che comporta il loro impiego.

In questo senso, il Regolamento distingue sistemi che presentano (a) rischi inaccettabili, (b) rischi di grado alto, (c) rischi legati prevalentemente alla trasparenza-conoscibilità (d) rischi minimi. Nell'ultima fase di negoziazione del regolamento, a seguito della diffusione di ChatGPT, è stata inoltre inserita un'ulteriore categoria di rischio, legata ai (e) sistemi per finalità generali (*General Purpose Artificial Intelligence*: GPAI).

Alcuni sistemi (sub a), presentando un chiaro pericolo per la salute, la sicurezza e i diritti fondamentali delle persone, sono considerati inaccettabili e banditi dal mercato unico a partire dal 2 febbraio 2025. Fra questi, a titolo esemplificativo, sono ricondotti: (i) sistemi che, attraverso l'impiego di tecniche subliminali, distorcono il comportamento di una persona, pregiudicandone la capacità di prendere decisioni consapevoli e comportando il rischio di un danno significativo; (ii) sistemi che sfruttano la vulnerabilità delle persone (per età o disabilità, ad esempio), distorcendone il comportamento, con il rischio di subire danni significativi; (iii) sistemi di classificazione delle persone basati sul comportamento o su caratteristiche personali che conducano ad un trattamento sfavorevole sproporzionato o in contesti scollegati da quelli di raccolta dei dati; (iv) sistemi che permettano di creare banche dati

---

<sup>11</sup> Maggiori dettagli nel considerando 25, AI Act.

di riconoscimento facciale mediante scraping indiscriminato di immagini facciali da internet; (v) sistemi di riconoscimento delle emozioni in ambito lavorativo o di istruzione (con eccezioni); (vi) alcuni sistemi di identificazione biometrica remota in tempo reale in spazi accessibili al pubblico<sup>12</sup>.

Altri sistemi (sub c) presentano rischi limitati alla possibilità che le persone che interagiscono con essi si confondano sulla loro natura (umana o artificiale), non riuscendo a distinguerli dalle persone. Il progresso tecnologico, infatti, ha fatto in modo che sempre più numerosi siano i sistemi talmente avanzati da superare il c.d. test di Turing e da essere considerati come portatori o imitatori di un comportamento connotato da intelligenza<sup>13</sup>. A fronte di questa eventualità, il Regolamento europeo ha posto in capo ai fornitori un dovere di progettazione tale che le persone siano informate del fatto di stare interagendo con una macchina e che i testi, gli audio, le immagini e i video creati da IA siano marcati e chiaramente riconoscibili come generati artificialmente (art. 50.1 e 50.2). Agli utilizzatori (*deployers*) di sistemi che creano contenuti *deep fake*, inoltre, il regolamento pone l'obbligo di rendere conoscibile la natura artificiale degli stessi (art. 50.4). Tali doveri entreranno in vigore a partire dall'agosto del 2026.

Nel caso in cui si tratti di sistemi che non pongono rischi di rilievo, il Regolamento non impone alcun obbligo, invitando comunque i fornitori ad adottare codici di condotta, che includano meccanismi di governance volti a promuovere l'applicazione volontaria dei requisiti di volta in volta funzionali a evitare qualsiasi rischio anche minimo (art. 95)<sup>14</sup>.

Invece, decisamente più complessa ed articolata risulta la disciplina riservata ai sistemi ad alto rischio (sub b). Va anzitutto menzionato come l'Unione Europea abbia deciso di dotare di un certo grado di flessibilità e adattabilità nel tempo i criteri in base ai quali individuare tali sistemi e la lista delle materie interessate. Tale approccio *future proof* è dovuto al rapido incedere delle innovazioni nel settore dell'IA che avrebbe condotto liste e criteri eccessivamente rigidi a diventare rapidamente obsoleti e anacronistici. Per questo, le indicazioni utili ad individuare quali siano i sistemi ad alto

---

<sup>12</sup> La lista completa dei sistemi vietati all'art. 5 dell'AI Act.

<sup>13</sup> Si tratta del test che permette di determinare, secondo Alan Turing, se una macchina abbia la capacità di esibire o imitare un comportamento intelligente: cfr. Alan M. Turing, "Computing machinery and intelligence", *Mind*, 49 (1950): 433-460.

<sup>14</sup> Maggiori dettagli nel considerando 165, AI Act.

rischio non sono presenti all'interno del Regolamento stesso, che prevede per la propria revisione procedure particolarmente complesse e dispendiose in termini di tempo, ma sono riportate all'interno di due allegati (*annexes I e III*), indicati dall'art. 6, i quali possono essere modificati ed aggiornati più agilmente dalla sola Commissione. L'allegato terzo, al riguardo, indica come sistemi ad alto rischio quelli impiegati a fini biometrici, in ambito di istruzione, formazione e lavoro, per l'accesso ai servizi essenziali, per le attività di contrasto (come poligrafi, strumenti di valutazione dell'affidabilità probatoria o del rischio di commissione di reato o di recidiva), in ambito di migrazione e controllo delle frontiere, nell'amministrazione della giustizia (strumenti di assistenza per la ricerca o interpretazione dei fatti o del diritto applicabile) e all'interno dei processi democratici (come quelli che potrebbero influenzare gli esiti elettorali). Tali sistemi dovranno conformarsi alle previsioni dell'AI Act entro l'agosto del 2026.

Nell'allegato primo, che sarà applicabile a partire dall'agosto 2027, sono invece caratterizzati come ad alto rischio i sistemi impiegati all'interno di una numerosa serie di apparecchiature già oggetto di disciplina europea, fra cui giocattoli, imbarcazioni, veicoli a motore e dispositivi medici. In riferimento all'individuazione dei sistemi ad alto rischio, infine, va ricordato come l'AI Act preveda una deroga riferita ai sistemi indicati nell'allegato terzo che non presentino un rischio significativo per la salute, la sicurezza o i diritti fondamentali delle persone, anche nel senso di non influenzare materialmente il risultato del processo decisionale. Nonostante l'art. 6.3 contenga alcune esemplificazioni dei casi coperti dall'eccezione (sistemi che eseguano un compito preparatorio o procedurale limitato, quelli destinati a migliorare il risultato di un'attività umana precedentemente completata, quelli non finalizzati a sostituire o influenzare una valutazione precedentemente completata senza un'adeguata revisione umana), tale disposizione pone problemi interpretativi complessi, potendo essere sfruttata strumentalmente al fine di evitare il rispetto degli impegnativi requisiti imposti ai sistemi ad alto rischio.

Una volta che un sistema sia ricondotto a tale categoria, decisione che peraltro compete allo stesso fornitore, l'AI Act prevede, infatti, il rispetto di una serie di condizioni non agevoli. Si tratta, anzitutto, della creazione di un sistema di gestione dei rischi, in grado di identificarli, valutarli e mitigarli (art. 9). In riferimento ai dataset utilizzati per l'addestramento, la convalida e la

prova, in secondo luogo, il Regolamento prevede la pertinenza, la sufficiente rappresentatività e, nei limiti del possibile, l'esattezza e la completezza, raccomandando l'utilizzo di proprietà statistiche appropriate (art. 10). In questo modo, si tenta di scongiurare in partenza la generazione di output scorretti o discriminatori. Oltre alla redazione della documentazione tecnica (art. 11), l'AI Act raccomanda la registrazione dei *log* per tutta la durata del ciclo di vita del sistema (art. 12), nel tentativo di garantire la tracciabilità del funzionamento dei sistemi ad alto rischio. In base all'art. 13 del Regolamento, inoltre, i fornitori sono destinatari di un obbligo teso a garantire che il funzionamento dei loro sistemi sia sufficientemente trasparente, in modo da permettere all'utilizzatore, anche sulla base di concise, complete, corrette e chiare informazioni d'uso, di interpretare gli output prodotti e di utilizzarli in modo appropriato. Tale previsione si collega all'art. 86 dell'AI Act che dispone il diritto della persona interessata da una decisione assunta con l'assistenza dei sistemi ad alto rischio elencati all'allegato terzo ad ottenere dall'utilizzatore spiegazioni chiare e significative sul ruolo assunto dall'IA e sui principali elementi della decisione adottata. Tali requisiti – come si vedrà *infra* – non sono di facile concretizzazione, visto soprattutto il problema della c.d. *Black Box*<sup>15</sup>.

Un ultimo requisito fondamentale per quanto riguarda tale categoria di sistemi si collega al principio del c.d. *human in the loop* o della sorveglianza umana (art. 14). L'AI Act, così, pone in capo ai fornitori l'obbligo di sviluppare sistemi la cui struttura permetta una supervisione umana, finalizzata alla mitigazione dei rischi per la salute, la sicurezza e i diritti fondamentali. A tale obbligo corrisponde, in capo all'utilizzatore, l'onere di contrastare la tendenza a fare eccessivo affidamento sulla correttezza dell'output (*automation bias*), e di giungere alla comprensione delle potenzialità e dei limiti del sistema, in modo da poterne individuare anomalie e disfunzioni e discostarsi da ogni output ritenuto scorretto o discriminatorio. Se quello della sorveglianza umana è uno dei cardini tanto del diritto quanto dell'etica dell'IA, va ricordato come la sua effettiva e concreta realizzazione dipenda da presupposti non sempre facili da riscontrare. Si tratta, da un lato, di una competenza tecnica diffusa che permetta agli utilizzatori di comprendere e interpretare il funzionamento di sistemi caratterizzati da un'estrema

---

<sup>15</sup> Frank Pasquale, *The Black Box Society*, New Haven: Harvard University Press, 2016.

complessità<sup>16</sup>; dall'altro, di una non facile assunzione di responsabilità personale che richiede tempo, impegno e forte motivazione individuali così come forme di supporto a livello strutturale<sup>17</sup>.

Molte altre sono le disposizioni che completano la struttura dell'AI Act (che conta, oltre a 180 considerando, 13 allegati e 113 articoli). Quanto illustrato, tuttavia, pare sufficiente a fornire il quadro e le premesse necessarie a svolgere un'adeguata trattazione della complessa applicabilità del principio di trasparenza (*rectius* di alcune sue declinazioni, riconducibili a spiegabilità, conoscibilità e tracciabilità) all'ambito delle attività di ricerca.

#### INTELLIGENZA ARTIFICIALE, RICERCA E TRASPARENZA. POSSIBILI CHIAVI DI LETTURA ALLA LUCE DELL'AI ACT

L'inquadramento normativo offerto dall'AI Act in materia di sviluppo e uso dei sistemi di IA si caratterizza, quindi, per una portata applicativa ampia e generale, tale da produrre effetti su numerosi settori e contesti di applicazione. Tra questi rientra anche l'ambito della ricerca, all'interno del quale l'IA, come si è già avuto modo di anticipare, può assumere rilevanza sia come oggetto, sia come strumento di questa attività. Infatti, nonostante il Legislatore europeo abbia predisposto una clausola volta a escludere l'applicazione delle disposizioni dell'AI Act ai sistemi e ai modelli sviluppati per sole finalità di ricerca e non si applica alle attività di ricerca, prova o sviluppo relative a sistemi o modelli prima della loro immissione sul mercato o messa in servizio, tale esenzione non deve considerarsi assoluta. Come indicato in precedenza, possono sussistere situazioni e circostanze in cui l'attuazione delle regole dell'AI Act si rende necessaria anche in relazione all'attività di ricerca, soprattutto nei casi in cui questa si svolga in condizioni reali e con il coinvolgimento di persone<sup>18</sup>.

<sup>16</sup> Al riguardo, l'art. 4 dell'AI Act prevede che fornitori e utilizzatori adottino misure adeguate a garantire al proprio personale un livello sufficiente di alfabetizzazione in materia di IA.

<sup>17</sup> Si potrebbe provocatoriamente affermare che in questo modo si richieda l'azione di una sorta di supereroe: cfr. Carlo Casonato, "Unlocking the Synergy: Artificial Intelligence and (old and new) Human Rights", *BioLaw Journal – Rivista di BioDiritto*, 3 (2023): 233-240.

<sup>18</sup> L'opportunità di applicare le regole dell'AI Act a questi contesti è evidenziata, oltre che dal considerando 25, anche da quanto stabilito dagli artt. 2, par. 8, e 60, Regolamento (UE) 2024/1689. A questo proposito cfr. Liane Colonna, "The AI Act's Research Exemption: A Mechanism for Regulatory Arbitrage?", in *YSEC Yearbook of Socio-Economic Constitutions 2023*, a cura di Eduardo Gill-Pedro, Andreas Moberg, Cham: Springer, 2024, 51 ss.

Tale rilievo pone di fronte alla necessità di comprendere quali disposizioni dell'AI Act possano assumere importanza in relazione a questo specifico contesto, con particolare riferimento a uno dei concetti che, sia dal punto di vista giuridico sia dalla prospettiva etica, si colloca come elemento cardine nello svolgimento delle attività di ricerca: il principio di 'trasparenza'.

Questo principio, che nella prospettiva dello Stato costituzionale di diritto si pone in un rapporto di stretta correlazione con i concetti di separazione dei poteri e di principio democratico<sup>19</sup>, trova fondamento nell'idea che la tutela delle persone e dei loro interessi sia assicurata anche grazie alla possibilità di rendere accessibili e di diffondere le informazioni che riguardano l'esercizio di una forma di potere, consentendone le necessarie limitazioni e modalità di controllo<sup>20</sup>.

Tale assunto si concretizza anche nella dimensione della ricerca dove, tradizionalmente, la condivisione dei dati costituisce una garanzia fondamentale sia per assicurare l'attendibilità, la riproducibilità e la verificabilità dei risultati prodotti, sia per tutelare i diritti e le libertà delle persone coinvolte come partecipanti in questa attività. Pure in questo caso, la conoscenza e la disponibilità delle informazioni risultano essere fattori fondamentali nel bilanciare situazioni di squilibrio di potere che possono emergere alla luce di diversi soggetti e dei molteplici interessi che in questo campo trovano affermazione e nel garantire il rispetto dei requisiti di integrità che devono permeare il campo della ricerca<sup>21</sup>.

Queste considerazioni circa la portata e gli obiettivi perseguiti nell'implementazione del principio di trasparenza assumono ancor più importanza in relazione a una tecnologia come l'IA. Che venga in gioco come oggetto o come strumento della ricerca, l'IA pone numerose questioni che sfidano i contenuti fondativi del concetto stesso di trasparenza e prospettano

---

<sup>19</sup> Norberto Bobbio, *Il futuro della democrazia*, Torino: Einaudi, 2013.

<sup>20</sup> Mark Fenster, "Seeing the State: the Transparency as Metaphor", *Administrative Law Review*, 62, 3 (2010): 617-672; Deirdre Curtin, Albert J. Meijer, "Does transparency strengthen legitimacy?", *Information Polity*, 11, 2 (2006): 109-122; Enrico Carloni, *Il paradigma della trasparenza. Amministrazioni, informazione, democrazia*, Bologna: Il Mulino, 2022, 61 ss.; Enzo Cheli, "Informazione, decisione politica, controllo sociale: spunti per un'analisi comparata", *Il diritto dell'informazione e dell'informatica*, 3 (198): 813-824.

<sup>21</sup> Lucia Busatta, "L'integrità della ricerca nel tessuto costituzionale: prime notazioni a partire dal contesto pandemico", *Rivista AIC*, 4 (2020): 389 ss.

profili tali da richiedere nuove chiavi di lettura<sup>22</sup>. In tale senso, l'analisi di quanto previsto dall'AI Act consente di leggere il significato di trasparenza alla luce delle esigenze poste dai sistemi di IA e di individuarne eventuali nuove declinazioni interpretative. Il ragionamento, tuttavia, va condotto alla luce del diverso inquadramento normativo offerto dal Regolamento europeo qualora l'IA sia oggetto o strumento della ricerca.

#### L'IA COME OGGETTO E COME STRUMENTO DELLA RICERCA: LE NUOVE DECLINAZIONI DEL PRINCIPIO DI TRASPARENZA DALL'AI ACT AI CODICI DI CONDOTTA

La prima ipotesi di lettura del principio di trasparenza trae origine dallo scenario in cui l'IA costituisce l'oggetto delle attività di ricerca, in una fase di sviluppo che si colloca idealmente in un momento immediatamente precedente all'immissione in commercio e alla messa in servizio del sistema di IA<sup>23</sup>. Da questa prospettiva, le disposizioni dell'AI Act applicabili e riconducibili al concetto di trasparenza sono molteplici e trasversali rispetto alla classificazione basata sul rischio adottata dal Regolamento, a dimostrazione dell'importanza riconosciuta a questo principio nel processo di regolamentazione dell'IA. Ciò che però rileva maggiormente sono le diverse declinazioni che vengono date al principio di trasparenza nel tentativo di elaborare soluzioni ai problemi posti dai sistemi di IA. Nello specifico, tre sono le principali linee di lettura che possono essere date al principio in esame alla luce di quanto stabilito dall'AI Act<sup>24</sup>.

---

<sup>22</sup> Erik Longo, Andrea Pin, "Oltre il costituzionalismo? Nuovi principi e regole costituzionali per l'era digitale", *Diritto pubblico comparato ed europeo*, 1 (2023): 103-116. A questo proposito si consenta il rinvio a Marta Fasan, *Intelligenza artificiale e costituzionalismo contemporaneo. Principi, diritti e modelli in prospettiva comparata*, Napoli: Editoriale Scientifica, 2024, 114 ss.

<sup>23</sup> Sul punto si veda quanto ribadito al considerando 25, Regolamento (UE) 2024/1689 e quando affermato in Amedeo Santosuosso, Giovanni Sartor, *Decidere con l'IA. Intelligenze artificiali e naturali del diritto*, Bologna: Il Mulino, 2024, 182 ss.; Marco Bassini, "Oggetto, campo di applicazione e ambito territoriale", in *La disciplina dell'intelligenza artificiale*, a cura di Oreste Pollicino et al., Milano: Giuffrè, 2025, 133 ss. È opportuno sottolineare come le regole introdotte dall'AI Act non si applichino ai sistemi di IA rilasciati con licenza libera e open source, a meno che questi non vengano immessi sul mercato come sistemi classificabili come proibiti, ad alto rischio o con problemi di trasparenza, secondo quanto previsto dall'art. 2, par. 12, Regolamento (UE) 2024/1689.

<sup>24</sup> La necessità di attribuire nuovi significati al principio di trasparenza era stata evidenziata anche nelle "Ethics Guidelines for Trustworthy AI", High Level Expert Group on Artificial Intelligence (2019). <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

La prima declinazione che può essere data al concetto di trasparenza in considerazione di quanto disposto dall'AI Act si pone in termini di 'spiegabilità' del sistema di IA. Con questo termine si fa riferimento, nello specifico, alla necessità di fornire spiegazioni sulle modalità di funzionamento tecnico del sistema intelligente e sulle motivazioni poste alla base dei risultati elaborati grazie all'apporto di questa tecnologia, così da assicurare tutela alle persone interessate dalle conseguenze e dagli effetti di decisioni adottate da e con l'IA<sup>25</sup>. In tal senso, il semplice accesso alle informazioni sulla logica decisionale e sui procedimenti tecnici seguiti dal sistema non è di per sé una misura sufficiente a fornire le garanzie tipiche della trasparenza, ma si richiede anche la capacità di comprendere questi profili e di saper dare una spiegazione in merito alle ragioni e agli obiettivi perseguiti dal sistema intelligente durante il suo funzionamento<sup>26</sup>. Solo grazie a interventi orientati ad assicurare la spiegabilità dell'IA è, infatti, possibile provare a realizzare una minimizzazione dei rischi derivanti dal già menzionato fenomeno della c.d. *Black Box* e dal suo conseguente impatto sulla vita delle persone.

L'importanza di questa declinazione trova, per l'appunto, conferma in alcuni contenuti dell'AI Act che si pongono proprio nella direzione di prevedere obblighi e requisiti funzionali a tale scopo. Così, sono riconducibili al concetto di 'trasparenza-spiegabilità' le disposizioni che per i sistemi 'ad alto rischio' prevedono una progettazione e uno sviluppo tali da assicurare che l'utilizzatore sia in grado di conoscere il funzionamento del sistema di IA e di interpretarne i risultati per impiegare la tecnologia in modo adeguato<sup>27</sup>. Queste previsioni sono, poi, ulteriormente rafforzate dall'obbligo di predisporre istruzioni per l'uso che permettano al c.d. *deployer* di comprendere le funzionalità offerte dall'IA e di impiegare il sistema adeguatamente secondo le sue finalità d'utilizzo<sup>28</sup>. L'esigenza di

---

<sup>25</sup> Luciano Floridi, *op. cit.*, 100 ss.; Karen Yeung, "Why Worry about Decision-Making by Machine?", in *Algorithmic Regulation*, a cura di Karen Yeung, Martin Lodge, Oxford: Oxford University Press, 2019, 28-29; Monica Palmirani, "Big Data e conoscenza", *Rivista di filosofia del diritto*, 1 (2020): 74 ss.; Elisa Spiller, "Il diritto di comprendere, il dovere di spiegare. Explainability e intelligenza artificiale costituzionalmente orientata", *BioLaw Journal – Rivista di BioDiritto*, 2 (2021): 419-432.

<sup>26</sup> Elettra Stradella, "AI, tecnologie innovative e produzione normativa: potenzialità e rischi", *DPCE online*, 3 (2020): 3361 ss.; Madalina Busuioc, Deirdre Curtin, Marco Almada, "Reclaiming transparency: contesting the logics of secrecy within the AI Act", *European Law Open*, 2, 1 (2023): 79-105.

<sup>27</sup> Art. 13, par. 1, Regolamento (UE) 2024/1689.

<sup>28</sup> Art. 13, par. 2, Regolamento (UE) 2024/1689.

fornire agli utenti le opportune spiegazioni per capire il funzionamento dell'IA e il significato dei suoi risultati si dimostra ancor più cruciale nelle situazioni in cui si richieda un intervento umano. Infatti, anche nell'attuazione delle misure di sorveglianza umana previste dall'AI la corretta comprensione del funzionamento del sistema e la capacità di interpretare i risultati da questo generati, fornendone una spiegazione, sono elementi essenziali per l'esercizio di un controllo attento sull'impiego di questa tecnologia<sup>29</sup>. È poi espressione di questa nuova declinazione del principio di trasparenza anche il riconoscimento del diritto alla spiegazione dei singoli processi decisionali. Nello specifico, l'AI Act prevede che le persone destinatarie di una decisione basata sui risultati prodotti da un sistema di IA, e da cui derivano effetti giuridici, hanno il diritto a ottenere dal *deployer* spiegazioni chiare e significative sul ruolo assunto dal sistema intelligente nel processo decisionale e sui principali elementi che contraddistinguono la decisione finale<sup>30</sup>.

La seconda declinazione attribuibile al principio di trasparenza si pone, invece, in termini di 'conoscibilità' del sistema di IA. Questo concetto può essere interpretato secondo due diverse chiavi di lettura, da cui discendono obblighi e requisiti diversi alla luce dell'AI Act. Nella prima, la nozione di conoscibilità si radica nella garanzia che siano rese note e condivise le informazioni riguardanti sia la natura artificiale del sistema utilizzato, sia la presenza dello stesso in procedimenti e interazioni che possano produrre effetti sulle persone<sup>31</sup>. Nella seconda chiave di lettura, invece, il concetto di conoscibilità si incardina sulla necessità di rendere note ai soggetti interessati le caratteristiche dell'IA, secondo un'accezione di trasparenza legata, anche in questo caso, alla conoscenza delle modalità di funzionamento di tale tecnologia<sup>32</sup>. Queste declinazioni interpretative fanno, quindi, riferimento alla necessità di prevedere idonee garanzie affinché le persone siano consapevoli dell'intervento di un sistema di IA, ne conoscano le funzionalità e che, in conseguenza di ciò, abbiano la possibilità di scegliere liberamente se avvalersi di questa tecnologia e dei suoi risultati.

---

<sup>29</sup> Art. 14, par. 4, lett. a) e c), Regolamento (UE) 2024/1689.

<sup>30</sup> Art. 86, Regolamento (UE) 2024/1689.

<sup>31</sup> Carlo Casonato, "Unlocking the Synergy: Artificial Intelligence and (old and new) Human Rights", *op. cit.*, 235 ss.

<sup>32</sup> Virginia Dignum, *Responsible Artificial Intelligence. How to Develop and Use AI in a Responsible Way*, Cham: Springer, 2019, 60-62; Thomas Wischmeyer, "Artificial Intelligence and Transparency: Opening the Black Box", in *Regulating Artificial Intelligence*, a cura di Thomas Wischmeyer, Timo Rademacher, Cham: Springer, 2020, 95.

Tali finalità sono confermate, anche in questo caso, dalle disposizioni dell'AI Act che intervengono ad arginare le conseguenze negative di una mancata autodeterminazione in merito all'impiego e agli effetti dell'IA e danno, quindi, attuazione al concetto di trasparenza-conoscibilità nelle sue due accezioni. Così, l'AI Act stabilisce esplicitamente l'obbligo di informare le persone destinatarie di una decisione adottata con l'intervento dell'IA dell'uso di un sistema classificato ad alto rischio e della loro soggezione ai risultati che ne derivano, assicurando che sia resa nota e conosciuta la presenza dell'IA nel processo decisionale<sup>33</sup>. Similmente, l'AI Act predispone garanzie di conoscibilità da applicare anche nei contesti di interazione diretta con il sistema di IA. In questo senso, si prevede che il sistema di IA sia progettato e sviluppato in modo tale che la persona, che interagisce direttamente con esso, sia consapevole di relazionarsi con un'IA e che i contenuti elaborati dai modelli di IA generativa abbiano natura artificiale<sup>34</sup>. Sono, poi, espressione di questa seconda declinazione di trasparenza anche le disposizioni indirizzate ad assicurare la conoscenza delle caratteristiche tecniche del sistema prima del suo effettivo impiego. Che si tratti di sistemi di IA o di modelli di IA per finalità generali, l'AI Act stabilisce specifici obblighi sulla condivisione delle informazioni e delle specifiche tecniche che riguardano le capacità, le proprietà e i limiti del sistema utilizzato<sup>35</sup>, anche prevedendo l'onere di redigere e di fornire la documentazione tecnica rilevante<sup>36</sup>, con l'obiettivo di rendere gli utilizzatori maggiormente consapevoli delle funzionalità del sistema e, quindi, di come impiegare la tecnologia in modo appropriato secondo le sue finalità d'uso<sup>37</sup>.

Infine, la terza declinazione data al principio di trasparenza nel quadro regolatorio fornito dall'AI Act si sostanzia nel concetto di 'tracciabilità'. Questo termine esprime la necessità che tutte le informazioni relative alla

---

<sup>33</sup> Art. 26, par. 11, Regolamento (UE) 2024/1689.

<sup>34</sup> Art. 50, Regolamento (UE) 2024/1689. Per un commento alla disposizione in esame e al suo contenuto normativo si veda Erik Longo, Federica Paolucci, "The Article 50 of the AI Act and the Transparency Obligations: the Model and its Limitations", in *La disciplina dell'intelligenza artificiale*, a cura di Oreste Pollicino et al., Milano: Giuffrè, 2025, 275 ss.

<sup>35</sup> Art. 13, par. 3, lett. b), Regolamento (UE) 2024/1689; Art. 53, par. 1, lett. b), Regolamento (UE) 2024/1689.

<sup>36</sup> Art. 11, Regolamento (UE) 2024/1689; Art. 53, par. 1, lett. a), Regolamento (UE) 2024/1689.

<sup>37</sup> Brent Mittelstadt et al., "The ethics of algorithms: Mapping the debate", *Big Data & Society*, 2 (2016): 6-7.

raccolta dei dati, alla loro classificazione, alla tipologia di algoritmo impiegato e riguardanti le modalità di adozione della decisione finale siano documentate correttamente, così da risultare tracciabili<sup>38</sup>. La previsione di regole e requisiti orientati a implementare una tale accezione di trasparenza trova ragion d'essere, ancora una volta, nel bisogno di mantenere il controllo su alcuni elementi e operazioni che contraddistinguono il funzionamento dell'IA per evitare che insorgano rischi durante il suo impiego. In tal senso, il concetto di 'trasparenza-tracciabilità' è funzionale all'individuazione di *bias*, errori, lacune e imprecisioni nei dataset utilizzati, alla conoscenza dei dati usati e delle modalità della loro analisi e all'identificazione di eventuali malfunzionamenti durante il processo decisionale, permettendo di limitare i danni che possano derivare da questa tecnologia e di tutelare le persone e i loro diritti<sup>39</sup>.

A questo scopo, l'AI Act introduce misure volte a monitorare la qualità dei dataset utilizzati, dal punto di vista dell'appropriatezza, della rappresentatività e della completezza delle informazioni usate, e a supervisionare le modalità di raccolta dei dati e le operazioni che riguardano il loro trattamento<sup>40</sup>. Inoltre, si prevede l'implementazione di meccanismi di registrazione dei *log* e, in generale, degli eventi più rilevanti del funzionamento dell'IA così da poter tenere traccia di tutti i processi realizzati, garantendo in questo modo la possibilità di individuare eventuali errori e problematiche e, se possibile, intervenire per porvi rimedio<sup>41</sup>.

Le esigenze di trasparenza sin qui descritte in riferimento all'attività di ricerca che abbia ad oggetto l'IA, si rilevano anche nel contesto in cui l'IA costituisca uno degli strumenti usati dai ricercatori per lo svolgimento delle proprie attività. In questo scenario, però, l'impostazione adottata dall'AI Act restituisce un quadro regolatorio totalmente diverso da quanto esaminato in precedenza. Il mancato inserimento di tale attività tra gli ambiti d'uso previsti

---

<sup>38</sup> Joanna J. Bryson, Andreas Theodorou, "How society can maintain human-centric artificial intelligence", in *Human-centered digitalizations and services*, a cura di Marja Toivonen, Evelina Saari, Singapore: Springer, 2019, 317 ss.

<sup>39</sup> Madalina Busuioc, Deirdre Curtin, Marco Almada, *op. cit.*, 86 ss.

<sup>40</sup> Art. 10, par. 2 e 3, Regolamento (UE) 2024/1689. Un obbligo simile è previsto anche per il *deployer* all'art. 26, par. 4, Regolamento (UE) 2024/1689.

<sup>41</sup> Art. 12, Regolamento (UE) 2024/1689. Una misura simile di registrazione dei log è prevista anche per i modelli di IA per finalità generali classificati con un rischio sistemico, secondo quanto previsto dall'art. 55, par. 1, lett. c), Regolamento (UE) 2024/1689. In generale, sugli specifici obblighi di trasparenza previsti per questi modelli di IA cfr. A. Santosuosso, G. Sartor, *op. cit.*, 183 ss.

dal Regolamento europeo implica, infatti, che il ricorso all'IA come strumento della ricerca non sia coperto dalla maggior parte delle regole previste dall'AI Act, ma si limiti alle disposizioni che abbiamo ricondotto al concetto di trasparenza-conoscibilità nell'interazione con il sistema.

Fatta eccezione per tali obblighi<sup>42</sup>, la disciplina di questo impiego dell'IA, e così anche l'eventuale declinazione del contenuto del principio di trasparenza, è affidata all'elaborazione di codici di condotta secondo quanto stabilito dall'AI Act<sup>43</sup>. Il Regolamento, infatti, prevede che si debba incoraggiare e agevolare l'elaborazione di codici di condotta per disciplinare le applicazioni dell'IA non classificate come proibite, ad alto rischio o con particolari problemi di trasparenza e che tali codici debbano promuovere l'applicazione volontaria dei requisiti previsti dall'AI Act per le altre tipologie di sistemi intelligenti. Secondo tale disposizione, quindi, il contenuto normativo dei codici di condotta dovrebbe ricalcare quanto già stabilito dall'AI Act (in particolare, al capo III, sezione 2) anche per quanto riguarda l'attuazione del principio di trasparenza, lasciando presupporre l'adozione di declinazioni interpretative simili a quelle già esaminate per i sistemi di IA che siano strumenti della ricerca<sup>44</sup>.

## CONCLUSIONI. L'AI ACT E L'ECOSISTEMA REGOLATORIO DELLA RICERCA

In entrambi i contesti che si sono analizzati, quello in cui l'IA è oggetto della ricerca e quello in cui essa opera come strumento, le letture date in termini di trasparenza-spiegabilità, trasparenza-conoscibilità e trasparenza-tracciabilità costituiscono importanti garanzie di fronte ai potenziali rischi legati all'uso di questa tecnologia, spesso connessi alla sua opacità riconducibile all'imperscrutabilità delle sue ragioni, alle sue abilità di mimesi dell'umano e alla complessità dei percorsi che ne determinano il funzionamento. In effetti, conoscere le ragioni alla base di un determinato risultato, essere consapevoli della natura di IA della tecnologia impiegata e avere cognizione dei dataset impiegati e delle operazioni svolte dal sistema

---

<sup>42</sup> Art. 50, Regolamento (UE) 2024/1689.

<sup>43</sup> Tale assunto deriva dal mancato inserimento del settore della ricerca quale ambito in cui l'uso dell'IA sia da considerarsi vietato o ad alto rischio, secondo l'inquadramento stabilito dagli artt. 5 e 6, Regolamento (UE) 2024/1689.

<sup>44</sup> Così previsto all'art. 95, Regolamento (UE) 2024/1689.

sono elementi essenziali per assicurare ai ricercatori e alle ricercatrici maggiore controllo sullo strumento utilizzato e, in definitiva, per garantire una ricerca di maggiore qualità<sup>45</sup>. Al contempo, essere a conoscenza del coinvolgimento di un sistema artificiale, così come comprenderne i meccanismi procedurali di base, e degli obiettivi di una ricerca sono condizioni essenziali per garantire una piena e consapevole adesione dei partecipanti alla stessa.

È evidente, però, che nel quadro normativo tracciato dall'AI Act permangono, allo stato dell'arte, alcune lacune e significative difficoltà interpretative che derivano, in particolare, dai confini applicativi del testo, dalla menzionata natura *product-centric* della regolamentazione e dall'approccio basato sul rischio. Di conseguenza, le norme relative alle tre declinazioni di trasparenza rinvenibili nell'AI Act troveranno diversa (o nessuna) applicazione a seconda che si tratti di una ricerca 'su' un sistema di IA o 'con' un sistema di IA, che questo sia o non sia ad alto rischio, che sia o non sia sviluppato e applicato per finalità di ricerca, che si agisca prima o dopo l'immissione del sistema in commercio, che si stia operando in condizioni reali, che dall'impiego del sistema di IA derivino decisioni che impattano sulle persone, che il sistema sia classificabile come *medical device*, che ci si interroghi sugli obblighi facenti capo agli sviluppatori dei sistemi (*providers*) o ai ricercatori che li utilizzano per lo svolgimento di attività di ricerca (*deployers*).

A fronte di questo quadro complesso di variabili non si può non segnalare un dato che pare essenziale tenere a mente: l'AI Act non opera in un vuoto, ma all'interno di un complesso ecosistema etico-normativo. Di questo si trova traccia anche nel considerando 25 dell'AI Act, il quale, dopo aver tentato un bilanciamento fra il bisogno di non ostacolare il progresso delle attività di ricerca e l'esigenza di mantenere punti fermi, almeno in riferimento ai sistemi ad alto rischio, rammenta come «(...) In ogni caso, qualsiasi attività di ricerca e sviluppo dovrebbe essere svolta conformemente alle norme etiche e professionali riconosciute nell'ambito della ricerca scientifica e dovrebbe essere condotta conformemente al diritto dell'Unione applicabile».

Le declinazioni della trasparenza delle quali si è trattato corrispondono, in effetti, ad esigenze tipiche del mondo della ricerca che da anni hanno

---

<sup>45</sup> Sulla possibilità che trovino applicazione le regole dell'AI Act quando l'IA sia utilizzata come strumento per condurre attività di ricerca cfr. Marco Bassini, *op. cit.*, 137-138; Hannah Ruschemeier, *AI as a challenge for legal regulation – the scope of application of the artificial intelligence act proposal*, ERA Forum, 23 (2023): 370 ss.

trovato riconoscimento in diversi documenti di *hard* e *soft law*. Si pensi alle corrispondenze fra la trasparenza-conoscibilità e i doveri di informazione gravanti sui ricercatori nei confronti del partecipante a una ricerca, la trasparenza-spiegabilità e l'esigenza di poter valutare il razionale di una ricerca scientifica e, ancora, la trasparenza-tracciabilità e il requisito di riproducibilità degli studi. Si tratta di doveri il cui rispetto è imposto dal complesso regolatorio che, a livello nazionale, sovranazionale e internazionale, garantisce il concetto di integrità di una ricerca e la tutela dei diritti delle persone coinvolte. L'AI Act, per come è formulato, non fornisce molti appigli normativi vincolati, ma contribuisce, come si è visto, a mettere in luce nuove declinazioni del concetto di trasparenza, delle quali chi opera nel mondo della ricerca, ampiamente inteso, dovrà tenere conto.

Se la trasparenza è un valore che nell'ambito della ricerca merita di essere preservato e se si vuole che l'IA sia effettivamente foriera di benefici in tale campo, è auspicabile, da un lato, che i riferimenti alle descritte declinazioni della trasparenza, come risposta alle opacità della tecnologia, trovino quanto prima traduzione all'interno dei menzionati codici di condotta e che, dall'altro, le regole dell'IA trovino la propria collocazione all'interno del più ampio ecosistema di regole etico-giuridiche che presidiano il campo della ricerca (si pensi, solo per fare alcuni esempi, alle norme del Regolamento sui trials clinici<sup>46</sup>, del Regolamento sui dispositivi medici<sup>47</sup> o, più latamente, anche del GDPR<sup>48</sup>).

Solo così sarà possibile definire con certezza le modalità di attuazione del principio di trasparenza nel contesto in esame, garantendo la qualità di una ricerca, sull'IA e con l'IA, che possa svolgersi all'interno di un virtuoso circuito di fiducia, risponda alla sua finalità di utilità sociale e risulti conforme ai canoni di integrità, nel rispetto dei diritti delle persone coinvolte.

---

<sup>46</sup> Regolamento (UE) n. 536/2014 del Parlamento europeo e del Consiglio, del 16 aprile 2014 sulla sperimentazione clinica di medicinali per uso umano e che abroga la direttiva 2001/20/CE.

<sup>47</sup> Regolamento (UE) 2017/745 del Parlamento europeo e del Consiglio, del 5 aprile 2017, relativo ai dispositivi medici, che modifica la direttiva 2001/83/CE, il Regolamento (CE) n. 178/2002 e il Regolamento (CE) n. 1223/2009 e che abroga le direttive 90/385/CEE e 93/42/CEE del Consiglio.

<sup>48</sup> Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati) (Testo rilevante ai fini del SEE).

