



FABIANA MAGNOLO*

LA COSTRUZIONE DI NUOVI DIRITTI FONDAMENTALI DINANZI ALLE SFIDE DELL'INTELLIGENZA ARTIFICIALE IN UN'OTTICA DI (RI)VALORIZZAZIONE DEL RUOLO UMANO

SOMMARIO: 1. Intelligenza artificiale e nuovi diritti. – 2. L'*AI disclosure* quale diritto dell'essere umano alla percezione circa l'alterità del suo interlocutore. – 2.1. I primi *exempla* di positivizzazione e la portata del Regolamento sull'intelligenza artificiale dell'Unione Europea. – 3. Trasparenza algoritmica e spiegabilità dei sistemi come diritto fondamentale. – 3.1. Il contenuto del diritto alla spiegazione attraverso i primi, timidi, riconoscimenti normativi e le prospettive aperte dall'*AI Act*. – 4. Il diritto ad uno "human in the loop". – 4.1. Il diritto ad un controllo umano sui sistemi di IA nel quadro giuridico attuale. – 4.2. Luci ed ombre dello "human oversight" nel Regolamento sull'IA dell'Unione Europea. – 5. Il possibile contenuto dei nuovi diritti fondamentali nell'ecosistema dell'intelligenza artificiale. – 6. Un solo "diritto" al cospetto di tante tecnologie. – 7. Riflessioni conclusive.

1. Intelligenza artificiale e nuovi diritti

Un nuovo diritto è una nuova istanza sociale, frutto di un mutamento nei costumi e riflesso di innovative esigenze di tutela, condivise da una generalità, più o meno ampia, di consociati. È l'essere cristallizzazione espressa di una coscienza collettiva cangiante l'elemento fondante il concetto di diritto quale posizione giuridica soggettiva attiva, vantata dal singolo e ritenuta meritevole di protezione proprio in quanto sentita parte integrante del patrimonio valoriale della comunità di riferimento di quell'individuo¹.

Ergo, la positivizzazione dei diritti non risulta decisiva ai fini della loro configurazione – ben potendo essi essere riconosciuti anche solo a livello giurisprudenziale –, ma logicamente postuma ed eventuale, perdendo la pretesa, una volta codificata, d'altro canto, il suo carattere di novità².

Tuttavia, l'operazione codificatrice assume un ruolo chiave nel percorso di identificazione delle nuove istanze sociali nella veste di diritti, poiché capace di elevare tali

* Dottoranda in Diritto internazionale presso l'Università del Salento.

¹ S. RODOTÀ, *Il diritto di avere diritti*, Roma-Bari, 2012, p. 7; M. LUCIANI, *Diritti fondamentali e nuove tecnologie*, in *Quaderni Costituzionali*, 2019, p. 8.

² P. GROSSI, *L'invenzione del diritto*, Roma-Bari, 2017, p. 112.

situazioni giuridiche, ove particolarmente meritevoli di tutela per la pregnanza collettiva rivestita ed il rango primario dei beni, *lato sensu*, coinvolti, allo *status* di veri e propri diritti non *tout court*, ma, *plus ultra*, fondamentali, i quali, *ex post* – ma, a tal uopo, non rileva – potrebbero trovare ingresso esplicito in carte e costituzioni.

Grande rilievo assume, pertanto, il riconoscimento di rango sub-costituzionale, il quale può fungere da indice ai fini dell'individuazione di nuove posizioni giuridiche³.

Quest'ultimo argomento avrà un ruolo centrale in questo lavoro, dedicato all'analisi dei possibili nuovi diritti fondamentali connessi all'era dell'intelligenza artificiale (IA)⁴. Si procederà, difatti, all'esame di alcune nascenti istanze sociali, generate dall'evoluzione tecnologica, ed in particolare dall'avvento dell'IA, fenomeni certamente incidenti sulla matrice sociale che si fa, così, portatrice di pretese di garanzia individuale dalle plausibili conseguenze di determinati utilizzi di tali strumenti. A riprova del loro essere diritti si analizzeranno anche le prime ipotesi di regolazione – di rango primario o secondario, a livello interno e sovranazionale – contenenti presidi a loro tutela. Tali possibili nuovi diritti, di cui si approfondirà nel prosieguo, includono: il diritto a una sorta di *AI disclosure*, ossia a conoscere la natura, umana o artificiale, dell'interlocutore coinvolto in concreto; il diritto ad una spiegazione dei risultati del sistema adoperato, ciò al fine di far rimanere in capo all'umano il controllo sulla tecnologia⁵, rappresentando questa istanza l'ultimo diritto di cui la scrivente, in tale sede, si occuperà.

Inoltre, sulla scia di un'osservazione proveniente da autorevole dottrina, che invita a procedere con particolare cautela al momento di riconoscere, in un determinato ordinamento, spazio ad un nuovo diritto⁶ – in quanto esso, al momento della sua concreta applicazione, potrebbe portare a restringere determinate prerogative altrui – sarà proprio la definizione del contenuto di tali situazioni giuridiche di recente apparizione a venire in rilievo, ovvero di quelle caratteristiche che le rendono veri e propri, nuovi, diritti fondamentali.

2. L'*AI disclosure* quale diritto dell'essere umano alla percezione circa l'alterità del suo interlocutore

L'intelligenza artificiale è ormai sempre più di comune impiego in vari – seppur specifici – aspetti della vita quotidiana, ove essa, per essere riconosciuta nella sua alterità, richiede uno sforzo particolare nel contesto pratico di riferimento, finendo, invero, per confondere gran parte degli utenti⁷.

³ A. RUGGERI, *Nuovi diritti e tecnologie emergenti*, in *Consulta Online*, 2020, p. 5.

⁴ Sul punto si veda S. RODOTÀ, *La costruzione dei diritti fondamentali nell'epoca digitale*, Roma, 2019; Y. MÈNY, *Human Rights in the Digital Age*, Londra, 2021; B. BIANCHINI, *The Future of Human Rights in the Era of Artificial Intelligence*, Oxford, 2022; S. PENASA, *Intelligenza artificiale e diritti fondamentali*, in *BioLaw Journal*, 2020, p. 13; A. SIMONCINI, *Costituzionalismo e nuove tecnologie*, in *Rivista AIC*, 2019, p. 4.

⁵ A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal – Rivista di BioDiritto*, 2019, p. 63 ss.; L. M. AZZENA, *L'algoritmo nella formazione della decisione amministrativa: l'esperienza italiana*, in *Revista Brasileira de Estudos Políticos*, 2021, pp. 503-537.

⁶ R. BIN, *Nuovi diritti e vecchie questioni*, in A.A.V.V., *Studi in onore di Luigi Costato*, III, Napoli, 2014; R. BIN, P. CHIARELLA, *Critica della teoria dei diritti. Conversazione con Roberto Bin*, in *Ordines – per un sapere interdisciplinare nelle istituzioni europee*, 2018, p. 327; A. PACE, *Problematica delle libertà costituzionali*, Padova, 2003.

⁷ Si tratta, prima di tutto, di sistemi utilizzati per automatizzare comunicazioni di breve o media lunghezza, ed in particolare *chatbot* e assistenti vocali, o per la produzione di contenuti c.d. *deepfake*, ove la distinguibilità della natura artificiale di tali tecnologie, da parte del pubblico, rispetto all'agire umano è spesso nulla; sul punto, D. K. CITRON, R. CHESNEY, *Deep fakes: a looming challenge for privacy, democracy, and national security*, in *California Law*

Di fronte a tali applicazioni tecnologiche, 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 lungimirante⁸. Elevare tale nuova posizione giuridica al rango di vero e proprio diritto fondamentale appare operazione giustificata in virtù dei valori coinvolti, in un sistema ordinamentale che si fonda – e tale deve rimanere – sulla centralità dell’essere umano⁹. Il diritto alla *disclosure*¹⁰ dell’intelligenza artificiale sarebbe posto a presidio della sfera della personalità individuale, in particolare, in termini di autocoscienza e di autodeterminazione nelle scelte, minacciate dal rischio di una silente reificazione dell’essere umano, quasi sopravfatto dall’uso di tecnologie delle quali neppure percepirebbe l’alterità rispetto a sé¹¹.

2.1. *I primi exempla di positivizzazione e la portata del Regolamento sull’intelligenza artificiale dell’Unione Europea*

A ben notare, già in alcuni ordinamenti, sono previsti obblighi – e non diritti espressi – di *disclosure* circa l’impiego dell’intelligenza artificiale. Tra gli atti a riguardo, di particolare pregnanza è il *Bolstering Online Transparency Act* californiano¹², il quale, segnatamente, vieta l’uso di *bot online* per finalità di vendita di beni e servizi o di propaganda elettorale, qualora non ne sia dichiarata la natura artificiale¹³. Il *B.O.T. Act*, quindi, riconosce, nel limitato ambito di riferimento, un vero e proprio obbligo di trasparenza, a presidio del quale, tuttavia, gli strumenti di tutela sono alquanto scarni, *i.e.* sanzioni amministrative comminate dai poteri pubblici, non attivabili, invero, dalla persona fisica in concreto interessata.

Inoltre, in capo ai *providers* di *social networks* e *internet services*, essendo meri intermediari in tali ipotesi, non sono posti obblighi specifici né di controllo, né di esclusione, dai propri servizi, degli operatori sanzionati per aver disatteso l’obbligo di *disclosure* circa l’uso di strumenti di comunicazione automatizzata¹⁴.

Altri esempi di parziale positivizzazione si rinvengono nella *Directive on automated decision-making* canadese¹⁵ e nel *GDPR* europeo¹⁶, che impongono oneri di trasparenza in materia di IA, qualora essa 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

Review, 2019, p. 1753 ss.; 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*, 2021, pp. 407-432; M. DURANTE, *Potere algoritmico e informazione*, in *Rivista di Filosofia del Diritto*, 2019, p. 71; L. FLORIDI, *Il verde e il blu. Idee naïf per un mondo digitale sostenibile*, Milano, 2020, p. 39.

⁸ E. CELESTE, *Digital Constitutionalism. The Double-Edged Sword of Fundamental Rights in the Digital Age*, Oxford, 2022, p. 155.

⁹ L. FLORIDI, *Il verde e il blu*, cit., p. 91.

¹⁰ Termine usato nella versione inglese del *GDPR* europeo.

¹¹ 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, pp. 257-321.

¹² *California Senate Bill 1001, Bolstering Online Transparency Act*, 28 settembre 2018, pienamente in vigore dal luglio 2019.

¹³ Nello specifico, sul punto, paragrafo n. 17941 del *Business and Professions Code* californiano (cui il *B.O.T. Act* ha aggiunto il *Chapter 6* alla *Part 3* della *Division 7*).

¹⁴ Sul punto, il paragrafo 17942 lett. c) del *Business and Professions Code*, anch’esso aggiunto dal *B.O.T. Act*.

¹⁵ *Government of Canada, Directive on automated decision-making*, 1 aprile 2019.

¹⁶ Reg. UE n. 679 del 2016; sul tema che ivi rileva, si veda: T. ZARSKY, *Incompatible: The GDPR in the Age of Big Data*, in *Seton Hall Law Review*, 2017, p. 1003.

a «[provide] 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»¹⁷.

D'altra parte, gli artt. 13.2 lett. f), 14.2 lett. g) e 15.1 lett. h) del Regolamento europeo, disciplinanti obblighi informativi nei confronti dell'interessato del trattamento, con riferimento al contesto concreto in cui si procede alla raccolta dei dati, impongono di fornire informazioni «sull'esistenza di un processo decisionale automatizzato, compresa la profilazione»¹⁸.

Appare evidente il limite comune in entrambe le previsioni normative, ovvero l'ambito di applicazione estremamente circoscritto. Difatti, sia la direttiva canadese sia il GDPR richiamano un dovere di dichiarare l'utilizzo di intelligenza artificiale solo nell'ipotesi di decisioni automatizzate (nel caso europeo, *in ultra*, solo se risultato in tema di trattamento di dati personali¹⁹).

A tali ipotesi di regolazione settoriale, inoltre, si affianca il Regolamento dell'Unione Europea sull'intelligenza artificiale del 2024²⁰, che sembra, però, adottare una prospettiva di più ampio raggio²¹. Nello specifico, l'art. 52 del testo prevede obblighi di trasparenza applicabili a determinati sistemi di IA, a prescindere dalla fascia di rischio di appartenenza²² (anche ove essi non siano ad alto rischio²³ e, *ergo*, non trovino applicazione gli stringenti requisiti tecnici di cui agli artt. 8 ss.). Il paragrafo 1 dell'articolo in esame stabilisce, invero, l'obbligo, in capo ai fornitori di strumenti di intelligenza artificiale destinati ad interagire con

¹⁷ Ai sensi dell'*Appendix C* della *Directive*, la tipologia di informazioni richiesta varia a seconda del livello di rischio – ovvero dell'impatto su «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» - del sistema automatizzato coinvolto (*i.e.*, all'*Appendix B*, quattro classi di rischio: minimo; moderato; alto; molto alto).

¹⁸ Tutti e tre gli articoli, invero, impongono l'obbligo summenzionato nei medesimi termini. 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 della disciplina privacy*, Milano, 2019, pp. 147-154 e 158-164; 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, pp. 441-459; Y. BATHAEY, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, in *Harvard Journal of Law & Technology*, 31, 2, 2018, pp. 889-938; A. RAI, *Explainable AI: from black box to glass box*, in *Journal of the Academy of Marketing Science*, 2020, 48, 1, pp. 137-141; A. ADADI, M. BERRADA, *Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)*, in *IEEE Access*, 6, 2018, pp. 5244-5245; S. WACHTER, B. MITTELSTADT, L. FLORIDI, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR*, in *International Data Privacy Law*, 2017, pp. 80-82.

¹⁹ *Ergo*, sono esclusi i dati non rientranti nella definizione legislativa di dato personale, che consistano in decisioni totalmente automatizzate, eccettuata, così, una vastissima gamma di sistemi di supporto alle decisioni degli esseri umani di enorme rilievo pratico, e spesso determinanti per la formulazione di queste ultime; sul punto, L. EDWARDS, M. VEALE, *Enslaving the algorithm: from a “right to an explanation” to a “right to better decisions”?*, in *IEEE Security & Privacy*, 2018, p. 3.

²⁰ Regolamento (UE) 2024/1689 (anche noto come *AI Act*) del Parlamento europeo e del Consiglio del 13 giugno 2024, entrato in vigore il 1 agosto 2024; per un approfondimento, si veda C. NOVELLI, *L'Artificial Intelligence Act Europeo: alcune questioni di implementazione*, in *federalismi.it*, 2024, p. 95 ss.; F. FERRI (a cura di), *L'Unione europea e la nuova disciplina sull'intelligenza artificiale: questioni e prospettive*, in *Rivista quaderni AISDUE*, 2024; A. BERTOLINI, *Intelligenza artificiale e regolazione giuridica: il modello europeo dell'AI Act*, in *Politica del Diritto*, 2023, p. 463; M. EBERS, S. NAVAS, *The EU Artificial Intelligence Act: Risk-Based Regulation of AI Systems*, in *Common Market Law Review*, 2023, p. 11.

²¹ Le altre norme dell'Unione Europea in materia di digitale, invece, non trattano direttamente il problema; nel *Digital Service Act* del 2022, ad esempio, il termine *deepfake* non compare mai.

²² Il Regolamento sull'IA suddivide le applicazioni dell'intelligenza artificiale in quattro classi di rischio (rischio inaccettabile, alto, limitato, minimo o nullo) sottponendo ciascuna – fatta eccezione per la prima categoria la quale comprende sistemi vietati nell'UE – di esse ad un differente regime normativo.

²³ Art. 6 del Regolamento.

esseri umani, di predisporre meccanismi che permettano la riconoscibilità degli stessi dall'esterno, «a meno che ciò non risulti evidente dalle circostanze e dal contesto di utilizzo». Il paragrafo 3 si occupa, invece, nello specifico dei contenuti *deep fake*, 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» debbano «rendere noto che il contenuto è stato generato o manipolato artificialmente». È interessante notare come l'*AI Act* sembri attribuire a tali garanzie pari livello rispetto a quello dei diritti fondamentali enunciati, ad esempio, nella Carta di Nizza²⁴ o, comunque, di altri interessi che assumono rilievo primario nelle operazioni di bilanciamento, *in primis* la sicurezza e l'ordine pubblico. Lo stesso paragrafo 1 dell'art. 52, difatti, chiarisce che la *disclosure* non riguarda i «sistemi di IA autorizzati dalla legge per accettare, prevenire, indagare e perseguire reati, a meno che tali sistemi non siano a disposizione del pubblico per segnalare un reato». Ed ancora, a conferma di tale approccio, nel secondo comma del paragrafo 3 del medesimo articolo, si indicano alcune eccezioni con riferimento ai contenuti *deep fake*. La norma, invero, prevede che l'obbligo non trovi applicazione qualora «l'uso [del sistema di IA con cui viene prodotto il contenuto *deep fake*] è autorizzato dalla legge per accettare, 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. Trasparenza algoritmica e spiegabilità dei sistemi come diritto fondamentale

Altra questione ivi di rilievo è la c.d. opacità dell'intelligenza artificiale, termine usato per indicare le difficoltà e gli ostacoli che l'essere umano incontra nel comprendere il funzionamento di tali sistemi, ovvero nel tradurne il *modus operandi* ed i risultati raggiunti in termini di spiegazione razionale. Tuttavia, alcune applicazioni dell'intelligenza artificiale – soprattutto le reti neurali²⁵ – presentano un certo grado di opacità quale caratteristica strutturale propria, aspetto che le rende non di facile comprensione nemmeno ai membri della comunità scientifica di riferimento²⁶.

²⁴ Carta dei diritti fondamentali dell'Unione europea, proclamata nel dicembre 2000.

²⁵ Appare opportuno, in tale sede, richiamare la distinzione tra intelligenza artificiale *model-based* e *data-driven* - si veda, *ex multis*, S. RUSSELL, P. NORVIG, *Artificial intelligence: a modern approach* (4a ed.), Hoboken (NJ), 2021, p. 47 ss. - ove la seconda è alla base del settore dell'apprendimento automatico, in particolare, delle reti neurali, e si basa sull'elaborazione di grandi moli di dati. È in relazione a tale caratteristica che si innesta il problema dell'opacità: più la rete è «profonda», ovvero costituita da numerosi strati di nodi e, *ergo*, particolarmente estesa e complessa, più le sue prestazioni risultano sì accurate ma, nel contempo, difficili – quasi impossibili – da comprendere in termini di funzionamento.

²⁶ G. VILONE, E. LONGO, *Explainable artificial intelligence: a systematic review*, 2020 e *Notions of explainability and evaluation approaches for explainable artificial intelligence*, in *Information Fusion*, 2021, pp. 89-106; B. GOODMAN, S. FLAXMAN, *European Union Regulations on Algorithmic Decision-Making and a “Right to Explanation”*, in *AI Magazine*, 2017, p. 52; A. MANTELERO, *From Safety to Fundamental Rights: Artificial Intelligence and EU Law*, in *Computer Law & Security Review*, 2021, p. 6.

3.1. Il contenuto del diritto alla spiegazione attraverso i primi, timidi, riconoscimenti normativi e le prospettive aperte dall'AI Act

In tema di possibile riconoscimento di un diritto alla spiegazione è utile menzionare alcune norme del GDPR, il cui ambito di riferimento è – come già ricordato – il trattamento di dati in relazione a decisioni totalmente automatizzate. L'art. 22, par. 1, difatti, vieta che un individuo possa essere soggetto ad una decisione di tal fatta quando ciò produca «effetti giuridici che riguardano [l'interessato] o incidono in modo analogo significativamente sulla sua persona». Il secondo paragrafo prevede alcune eccezioni al divieto, stabilendo che siffatti trattamenti siano possibili ove necessari per la conclusione o l'esecuzione di un contratto tra titolare ed interessato, oppure si basino sul consenso di quest'ultimo, o ancora siano esplicitamente autorizzati dal diritto dell'Unione Europea o di uno Stato membro cui è soggetto il titolare del trattamento²⁷. L'ultimo paragrafo pone, però, dei limiti alle predette eccezioni nell'ipotesi in cui il trattamento abbia ad oggetto dati c.d. particolari, per i quali sono previsti presidi specifici proprio in ragione della loro intrinseca “sensibilità”²⁸. Ivi, le uniche basi giuridiche legittimanti sono il consenso esplicito dell'interessato (art. 9 lett. a) o il perseguimento di motivi di interesse pubblico rilevante sulla base del diritto dell'Unione o degli Stati membri (art. 9 lett. g). L'art. 22, inoltre, impone sempre, ove la decisione automatizzata sia consentita, l'adozione di misure concrete idonee ad una effettiva «tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato», precisando, per giunta, che, nei casi di trattamento necessario alla conclusione o all'esecuzione di un contratto, o basato sul consenso dell'interessato, esse debbano 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»²⁹.

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 circa l'esistenza del trattamento automatizzato e, in tale circostanza, devono essergli fornite «informazioni significative sulla logica utilizzata»³⁰. A ciò si aggiunge il Considerando 71 il quale indica alcune garanzie che è opportuno adottare ove sia lecito l'uso di una decisione automatizzata ai sensi dell'art. 22, tra le quali: «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».

Fin qui, a ben vedere, la sola norma a menzionare esplicitamente – seppur sempre in un contesto circoscritto – un diritto alla spiegazione è il Considerando 71, disposizione che, però, non ha valore precettivo, ma mera funzione di orientare l'interpretazione degli articoli del Regolamento³¹.

²⁷ L'ampiezza di tali eccezioni è stata criticata, cfr. ad es. A. SIMONCINI, *L'algoritmo incostituzionale*, cit., p. 79.

²⁸ A delimitarne il perimetro è l'art. 9 par. 1 che ne vieta, in generale, il trattamento; il paragrafo successivo elenca le eccezioni al divieto.

²⁹ Sul punto, F. PIZZETTI, *Decisioni automatizzate, algoritmi e tutele giuridiche*, in *Federalismi.it*, 2020, p. 14.

³⁰ 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 di accesso ai dati trattati.

³¹ Sul punto, anche le Linee Guida del febbraio 2018 dell'ARTICLE 29 DATA PROTECTION WORKING PARTY, poi ratificate dalla European Data Protection Board, una volta entrato in vigore il Regolamento, pag. 25; S. WACHTER, B. MITTELSTADT, L. FLORIDI, *Why a Right to Explanation*, cit., p. 95.

Intra scenario europeo, appaiono, inoltre, meritevoli d'analisi le norme di diritto interno francese, ove, nel 2016, con la *Loi n. 2016-1321 du 7 octobre 2016 pour une République numérique*, sono state introdotte alcune garanzie relative alla facoltà, in capo all'individuo, di ricevere una spiegazione circa decisioni amministrative assunte, anche non in via esclusiva, tramite algoritmi³².

In particolare, l'art. 4 della legge in questione 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 recita: «une décision individuelle prise sur le fondement d'un traitement algorithmique comporte une mention explicite en informant l'intéressé [...]. La "mention", specificamente, deve indicare, almeno, «la finalité poursuivie par le traitement algorithmique» e il 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»³³.

Secondo l'art. L. 311-3-1-2, tale "communication" deve, a sua volta, contenere informazioni il più possibile complete³⁴, ovvero fornire «sous une forme intelligible» notizie 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»³⁵.

D'altro canto, è bene ricordare che la Francia, è parte di quella minoranza di Paesi europei che hanno individuato, nel proprio ordinamento, spazi di legittimità, ulteriori rispetto a quelli di cui all'art. 22 GDPR, circa trattamenti consistenti in decisioni *in toto* automatizzate, in adempimento a quanto consentito dal secondo paragrafo, lett. b), della norma, con l'avvertimento di prevedere, in tali ipotesi, «misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato»³⁶.

L'art. 47 della *Loi informatique*³⁷ esclude, in *primis*, che una decisione totalmente automatizzata possa applicarsi in sede giudiziaria al fine di valutare il comportamento o la personalità di un individuo e specifica, in *secundis*, che una decisione di tal fatta debba essere formulata nel rispetto dei criteri summenzionati introdotti con la *Loi n. 2016-1321 du 7 octobre 2016 pour une République numérique*, ed impone, nel contempo, 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». Esso, per giunta, dispone che qualsiasi trattamento consistente in una decisione automatizzata possa avvenire solo in osservanza delle eccezioni di cui all'art. 22 GDPR, richiedendo, come garanzia addizionale, che ivi «des règles définissant le traitement ainsi que les principales caractéristiques de sa mise en œuvre

³² *Loi n. 2016-1321 du 7 octobre 2016 pour une République numérique*, *Journal officiel de la République française* 235, 8 ottobre 2016.

³³ Art. L. 311-3-1, introdotto, insieme al successivo articolo, con il *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.

³⁴ G. SARTOR, *L'amministrazione algoritmica*, in *Rivista Trimestrale di Diritto Pubblico*, 2021, p. 306.

³⁵ Si veda l'art. L. 311-3-1-2.

³⁶ Art. 22 par. 2 lett. b) del GDPR.

³⁷ L'intervento è stato attuato con la *Loi n. 2018-493 du 20 juin 2018 relative à la protection des données personnelles*, che 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*.

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»³⁸.

Extra contesto europeo, notabile *exemplum* verso una possibile positivizzazione del diritto alla spiegabilità è rappresentato dalla *Directive on automated decision-making* canadese, il cui par. 6.2.3 impone alle pubbliche amministrazioni che utilizzino algoritmi di fornire «a meaningful explanation to affected individuals of how and why the decision was made as prescribed in Appendix C»³⁹. Tale appendice prevede caratteristiche di spiegazione differenti a seconda del rischio connesso alla tecnologia impiegata⁴⁰. Invero, in caso di «little or no impact» (*Level I*), la spiegazione può essere semplificata, *i.e.* consistente in informazioni sulle comuni modalità di funzionamento del sistema, senza precipui collegamenti con la decisione adottata in concreto⁴¹. Invece, per ogni altra decisione – e, *ergo*, con riferimento ai restanti livelli di rischio – l'autorità che adopera gli algoritmi deve garantire al soggetto interessato una spiegazione specifica ove il risultato sia il rifiuto di un beneficio o di un servizio, o, in ogni caso, sia rinvenibile l'esercizio di un potere (nel testo «regulatory action»)⁴². Tuttavia, al pari delle norme *intra* UE in materia analoga – con le quali la direttiva canadese presenta punti di contatto anche in riferimento a scelte terminologiche, ad esempio l'uso del termine *meaningful*, quale carattere qualificante le spiegazioni, che nel GDPR è «significative» – la norma non individua in maniera dettagliata, o simil tale, quali informazioni debbano essere oggetto di spiegazione al fine di adempiere effettivamente all'obbligo imposto dalla direttiva.

Dalla disamina in tema di norme in materia di spiegabilità dell'intelligenza artificiale, emerge come esse in realtà non assicurino una piena tutela ad una posizione giuridica che, secondo la prospettiva adottata in questo lavoro, si eleva al rango di diritto fondamentale (*i.e.* diritto ad una spiegazione). A tal uopo, appare interessante fare menzione del Regolamento dell'Unione Europea del 2024 in materia di intelligenza artificiale. L'art. 13 prevede che i sistemi di IA ad alto rischio (il cui ambito è comunque molto ampio)⁴³ 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 ed 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*]»⁴⁴. Come appare

³⁸ Art. 47 della *Loi informatique et libertés*.

³⁹ La norma è compresa tra gli obblighi di trasparenza di varia natura previsti dal paragrafo 6.2, per l'appunto rubricato *Transparency*, che impone, *ex multis*, 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, anche esaminandone le singole componenti, qualora siano utilizzati strumenti coperti da segreto industriale (par. 6.2.4- 5).

⁴⁰ Come già riportato nella nota n. 10, l'*Appendix B* distingue quattro classi di rischio; R. CALO, *Artificial Intelligence Policy: A Primer and Roadmap*, in *UC Davis Law Review*, 2017, p. 420.

⁴¹ L'*Appendix C* nel delineare l'“*explanation requirement*” per il *Level I*, 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».

⁴² 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».

⁴³ Art. 6 del Regolamento.

⁴⁴ Così recita il primo paragrafo dell'articolo, mentre quelli successivi sono dedicati alle informazioni che devono essere fornite ove usate tecnologie ad alto rischio.

evidente, la norma ha natura fortemente tecnica, tanto che i requisiti di trasparenza da essa contemplati rappresentano un ausilio all'attività dell'utente del sistema ed uno strumento per il rispetto degli obblighi posti dal Regolamento al fornitore della tecnologia, senza tenere in conto il destinatario degli effetti degli *output* di tali applicazioni, che potrebbe essere persona diversa dall'utente. Tuttavia, stanti l'indubbio campo d'applicazione più ampio ed una formulazione del preceitto meno vaga, esso potrebbe rappresentare un valido passo in avanti nel dare adeguato peso giuridico ad interessi strettamente connessi alla spiegabilità dei sistemi. Dunque, nonostante nel Regolamento sull'IA non si faccia ricorso ad espressioni quali "diritto alla spiegazione", come, invece, avviene nel Cons. 71 del Regolamento in materia di dati personali, esso sembra poter maggiormente rappresentare indice di un'attenzione peculiare, da parte dell'ordinamento europeo, circa la garanzia di una pretesa dell'individuo alla spiegazione intesa quale vero e proprio diritto fondamentale.

4. *Il diritto ad uno "human in the loop"*

L'automazione, tramite l'intelligenza artificiale, di un numero sempre crescente di attività pone un interrogativo circa il ruolo residuo in capo all'essere umano⁴⁵. In un tale scenario, il riconoscimento di un certo margine di controllo e di eventuale possibilità d'intervento potrebbe configurarsi quale presidio nei confronti delle conseguenze, scaturenti da processi *in toto* automatizzati, potenzialmente lesive di diritti fondamentali⁴⁶.

Le modalità concrete di attuazione di una garanzia antropocentrica di tal fatta sono, a ben vedere, variabili e in stretta connessione con la tipologia di tecnologia di rilievo, aspetto che rende l'operazione affatto semplice.

Invero, esiste un indubbio nesso tra la possibilità di ricavare una spiegazione dalle operazioni del sistema e la sua sottoponibilità ad un controllo, più o meno esteso, da parte dell'essere umano⁴⁷.

4.1. *Il diritto ad un controllo umano sui sistemi di IA nel quadro giuridico attuale*

I principali tentativi di giuridicizzazione circa un certo spazio di controllo umano sulle nuove tecnologie si rinvengono, ancora una volta, nel GDPR e nella *Directive on automated*

⁴⁵ 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, pp. 364-381; F.M. ZANZOTTO, *Viewpoint: Human-in-the-loop Artificial Intelligence*, in *Journal of Artificial Intelligence Research*, 64, 2019, pp. 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, pp. 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; 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, pp. 944-965.

⁴⁶ Sul punto, ad esempio, C. GRIECO, *Intelligenza artificiale e diritti umani nel diritto internazionale e dell'Unione europea. Alla ricerca di un delicato equilibrio*, in *Ordine internazionale e diritti umani*, 2022, pp. 787-792; B. BRAUNSCHWEIG, M. GHALLAB, *Reflections on Artificial Intelligence for Humanity*, Berlino, 2021; A. ODDENINO, *Decisioni algoritmiche e prospettive internazionali di valorizzazione dell'intervento umano*, in *DPCE-online*, 2020, pp. 199-217.

⁴⁷ Tale legame è rinvenibile anche da una lettura delle Linee guida etiche del 2018 dal Gruppo di Esperti di Alto Livello sull'IA, p. 18, che individua quattro scenari di controllo, partendo da una totale assenza di sorveglianza umana (*human out of the loop*), passando per un margine maggiore di partecipazione in capo alla persona fisica (*human on the loop* e *human in the loop*), fino a giungere ad un controllo umano totale sul sistema (*human in command*).

decision-making canadese. Per quanto concerne il GDPR rileva, in particolare, il già menzionato art. 22, ove per due delle eccezioni al divieto enunciato dalla norma⁴⁸ – *i.e.* la conclusione o esecuzione di un accordo contrattuale ed il trattamento basato sul consenso – il par. 3 prevede, tra le misure a tutela dei «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»⁴⁹.

Nel garantire la possibilità di richiedere un intervento imputabile ad una persona fisica anche in ipotesi in cui sia eccezionalmente ammessa la decisione automatizzata, l'art. 22 sembra voler riconoscere il ruolo centrale dell'essere umano, assicurando un suo controllo effettivo, e non meramente potenziale, controllo che potrà assumere, come visto in precedenza, forme variabili a seconda delle caratteristiche della tecnologia coinvolta e della sua, ove sussistente, opacità.

Inoltre, l'art. 22 par. 3 riconosce all'interessato il diritto di esprimere la propria opinione e di contestare la decisione che lo riguarda, posizione giuridica questa da doversi ricondurre sempre all'area del riconoscimento di un controllo umano sul sistema, interpretandola come possibilità di sottoporre la decisione in questione ad una revisione operata dall'uomo o meglio a cura dello stesso titolare del trattamento⁵⁰, sul quale grava proprio un onere in tal senso.

Spostando lo sguardo *extra* UE, la summenzionata *Directive on automated decision-making* canadese impone, al paragrafo 6.3.9, di «ensuring that an Automated Decision System allows for human intervention, when appropriate, as per Appendix C»⁵¹. L'*Appendix C* precisa poi, appunto, i confini di tale *human intervention*, in funzione dei vari livelli di rischio associati al sistema e definiti, come già visto, dall'*Appendix B*. Per le due classi di rischio più basse non sono previste garanzie particolari e risultano leciti, *ergo*, processi decisionali privi di un intervento umano diretto. Qualora il rischio appaia alto o molto alto (*Levels III-IV* dell'*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»⁵².

4.2. Luci ed ombre dello “human oversight” nel Regolamento sull'IA dell'Unione Europea

A questo punto, appare utile soffermarsi sulle norme riguardanti i requisiti tecnici volti a garantire la presenza di un controllo umano previste dal già citato Regolamento europeo in materia di intelligenza artificiale.

⁴⁸ Ovvero divieto di sottoporre l'individuo ad 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».

⁴⁹ A. MANTELERO, *From Safety to Fundamental Rights*, cit., p. 9.

⁵⁰ Il legame tra l'elemento dell'intervento umano e la possibilità di contestare la decisione, così come 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.

⁵¹ Il paragrafo riportato, significativamente, apre la sezione della Direttiva rubricata *Ensuring human intervention*.

⁵² U. PAGALLO, *Algorithmic Decisions and the Human-in-the-Loop Principle*, in *Law, Innovation and Technology*, 2018, p. 139.

A rilevare è, in particolare, l'art. 14 del testo, rubricato *human oversight* da applicarsi a tutte le tecnologie considerate ad alto rischio⁵³. La norma, al paragrafo 2, esplicita le funzioni della sorveglianza umana⁵⁴ e ne sottolinea lo stretto legame con una serie di diritti fondamentali, affermando che essa «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, ed in particolare il paragrafo 4, dettano, invece, stringenti adempimenti tecnici di applicazione necessari affinché possa sussistere un adeguato livello di *human oversight*. È previsto, invero, l'obbligo di mettere in atto misure che rendano noto all'utente il rischio circa un eccessivo affidamento nei confronti della tecnologia e di garanzia della possibilità di fermare il funzionamento del sistema ed incidere sui risultati (individuando, in tale maniera, prerogative che costituiscono il fulcro del diritto al controllo umano).

Altre parti della norma, però, impongono requisiti più severi, strettamente legati al grado di trasparenza del sistema. Il menzionato art. 14 par. 4, ad esempio, prevede anche 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 rimandino allo stato dell'arte in ambito tecnologico – *ad exemplum* nel far riferimento ai «metodi di interpretazione disponibili» – esse sembrano richiedere, se *stricto sensu* interpretate, un livello di controllo in capo all'essere umano⁵⁵ che, in alcuni casi, potrebbe risultare di difficile attuazione⁵⁶ in relazione soprattutto a tecnologie ostiche da interpretare, come alcune applicazioni delle reti neurali⁵⁷.

5. Il possibile contenuto dei nuovi diritti fondamentali nell'ecosistema dell'intelligenza artificiale

Il primo dei nuovi diritti fondamentali, come *supra* enunciato, che ivi si propone di introdurre nell'ordinamento della c.d. «società algoritmica»⁵⁸, è il diritto alla riconoscibilità dell'alterità artificiale, *i.e.* diritto alla *AI disclosure*. Esso si configura, invero, quale proiezione diretta del principio, dal certo carattere personalista, della dignità umana⁵⁹, nella misura in cui garantisce all'individuo la possibilità di sapere se l'interazione nella quale è coinvolto avvenga con un soggetto a sé similare ovvero con un sistema artificiale dotato di autonomia funzionale.

Tale diritto non risponde a una mera esigenza informativa, ma incarna, a ben vedere, una più profonda istanza di tutela inerente all'autodeterminazione cognitiva e relazionale del soggetto. Ignorare la natura artificiale dell'interlocutore espone, difatti, l'individuo a «forme

⁵³ Si rimanda all'art. 14 dell'*AI Act*.

⁵⁴ *Human oversight* così tradotto nella versione in italiano.

⁵⁵ L'intensità dello *human oversight* qui definito sembrerebbe accostabile, utilizzando la categorizzazione delle Linee Guida etiche del Gruppo di Esperti di alto livello sull'IA, allo *human in the loop*.

⁵⁶ E. PALMERINI ET AL., *Human Oversight of Automated Decision-Making in the EU Legal Framework*, in *European Journal of Risk Regulation*, 2021, p. 14; A. BERTOLINI, *Intelligenza artificiale e regolazione giuridica*, cit., p. 478.

⁵⁷ Per un inquadramento circa tali sistemi si veda la nota n. 18.

⁵⁸ L. FLORIDI, *The Ethics of Artificial Intelligence*, Oxford, 2022.

⁵⁹ S. RODOTÀ, *Il diritto di avere diritti*, cit., p. 189; L. FLORIDI, *Il verde e il blu*, cit., p. 94.

subdole di manipolazione, influenza occulta e asimmetria cognitiva strutturale, che si pongono in frontale contrasto con i principi di libertà morale e di egualanza sostanziale»⁶⁰.

La *disclosure* dell'identità artificiale dell'interlocutore deve essere, *in primis*, chiara, tempestiva ed accessibile, e, *in secundis*, deve riguardare non soltanto la natura intrinseca del sistema, ma anche – e forse soprattutto – le sue finalità generali, il possibile – determinato in via presuntiva - impatto sui diritti individuali nonché i rimedi – previamente definiti, insieme ad un sistema prefissato, almeno in termini di principi generali, di attribuzione della relativa responsabilità - esperibili in caso di lesione degli stessi. In tal senso, la *disclosure* si configura come presupposto ontologico dell'esercizio di ogni altro diritto algoritmico: senza consapevolezza dell'interazione con una “macchina”, non vi può essere né spiegazione significativa, né controllo umano effettivo, né autentica possibilità di tutela giurisdizionale.

Il modello ivi delineato, invero, si colloca in linea di continuità critica con l'art. 52 dell'*AI Act* europeo - il quale introduce obblighi di trasparenza per i sistemi che interagiscono con gli esseri umani – distaccandosene, tuttavia, in senso garantista, mirando ad elevare, in un'inversione di prospettiva, tali obblighi al rango di diritto soggettivo fondamentale, concretamente azionabile e non quale mero precetto formale regolatorio⁶¹.

Strettamente connesso al predetto diritto dell'individuo alla percezione circa l'alterità del suo interlocutore è il diritto ad una spiegazione significativa delle decisioni algoritmiche, il quale, invero, costituisce uno degli snodi più complessi nella costruzione di un apparato unitario di diritti configurabili nell'era dell'intelligenza artificiale.

Tale diritto, così come in tale sede si intende delineare, si pone quale evoluzione del contenuto dell'art. 22 del GDPR - cristallizzante una forma embrionale di tutela contro le decisioni interamente automatizzate basate su dati personali - la cui portata, però, si è rivelata strutturalmente insufficiente, sia per i limiti soggettivi della previsione (invero, avente riferimento precipuo al trattamento dei dati personali), sia per ambiguità interpretative circa l'effettiva esistenza di un vero e proprio «*right to explanation*»⁶².

Il diritto ad una spiegazione ivi tratteggiato assume, invece, una dimensione graduata e non più settoriale, svincolata dal solo perimetro della protezione dei dati personali ed ancorata al criterio dell'incidenza sostanziale sui diritti – “tradizionali” - della persona⁶³. L'*explanation* non è concepita come un assoluto uniforme, bensì come un obbligo – cui corrisponde specularmente l'annesso diritto - modulato in relazione al rischio attribuibile alla tecnologia coinvolta nel caso di specie: *ergo*, si avrà un'informazione generale per i sistemi a basso rischio; una spiegazione operativa per quelli a rischio medio; una spiegazione individualizzata, eventualmente controfattuale, per i sistemi ad alto rischio.

Tale strutturazione risponde ad un'esigenza di realismo epistemologico, posto che molti sistemi di apprendimento automatico – in particolare, come predetto, quelli fondati su reti neurali profonde – sono caratterizzati da una opacità strutturale che rende impossibile una spiegazione pienamente causale in senso tradizionale⁶⁴. Il diritto ad una spiegazione deve, dunque, essere inteso come diritto ad un'informazione comprensibile e funzionalmente utile alla tutela della propria sfera personale, avente pregnanza e meritevolezza giuridiche, e non come, d'altronde, pretesa di totale trasparenza del codice sorgente.

⁶⁰ A. MANTELERO, *Artificial Intelligence and Fundamental Rights*, in *Computer Law & Security Review*, 2020.

⁶¹ M. VEALE, F.Z. BORGESIUS, *Demystifying the Right to Explanation*, in *Computer Law Review International*, 2018; G. SARTOR, *Human Rights and Information Technologies*, Oxford, 2015.

⁶² M. VEALE, F.Z. BORGESIUS, *Demystifying the Right to Explanation*, cit.

⁶³ M. J. TADDEO, L. FLORIDI, *Explainable AI as a Human Right*, in *Philosophy & Technology*, 2022, p. 7.

⁶⁴ B. MITTELSTADT ET AL., *The Ethics of Algorithms: Mapping the Debate*, in *Big Data & Society*, 2016.

Il terzo pilastro della proposta normativa, in seno a tale lavoro profilata, è rappresentato dal diritto al controllo umano effettivo, che segna il definitivo superamento dell'idea – ancora presente, come innanzi visto, in parte, *ad exemplum*, della regolamentazione europea – di un controllo umano meramente simbolico e formale.

Nel discorso giuridico-normativo *supra* trattato, il riferimento allo *human oversight* sembra configurarsi più come una sorta di clausola di stile, spesso svuotata di reale sostanza normativa. La previsione di un diritto soggettivo al controllo umano impone, invece, che l'intervento dell'individuo non sia ridotto ad una mera ratifica automatica dell'*output* algoritmico, ma si configuri come atto decisionale autonomo, informato e giuridicamente responsabile⁶⁵.

Il controllo umano, pertanto, per essere tale, deve soddisfare tre requisiti essenziali: l'accesso ad informazioni tecnicamente comprensibili, il possesso di un reale potere di modifica della decisione, nonché l'assunzione di una piena responsabilità giuridica dell'esito finale. In assenza anche di uno solo di tali elementi, il controllo degrada a mera finzione procedurale.

Tale impostazione si pone in linea con le più avanzate riflessioni etico-giuridiche sulla necessità di preservare una “ultima istanza umana” nelle decisioni che incidono su diritti fondamentali, in particolare nei settori della giustizia, della sanità, del credito e della sicurezza⁶⁶. Essa si distacca, *ergo*, dall'approccio prevalentemente tecnico adottato dall'*AI Act*, il quale si concentra, invero, su obblighi gravanti in capo a fornitori e soggetti terzi rispetto all'individuo titolare specifico dell'istanza, senza tradurli in un vero e proprio diritto fondamentale azionabile dal singolo.

Il diritto al controllo umano effettivo assurge, in tale maniera, a presidio ultimo della responsabilità giuridica nell'ecosistema algoritmico, impedendo che la catena delle decisioni si dissolva nell'irresponsabilità sistemica tipica dei processi automatizzati complessi.

6. Un solo “diritto” al cospetto di tante tecnologie

Come evidenziato nel corso del lavoro, allo stato dell'arte la regolazione dell'intelligenza artificiale in termini di riconoscimento ed incidenza sui – “nuovi” e “vecchi” – diritti fondamentali è costituita da sporadiche nonché sommarie ipotesi normative⁶⁷, ove la principale, in termini di distintività e pregnanza per ciò che ai fini della presente disamina rileva, delle quali è il menzionato Regolamento dell'UE sull'IA. La maggioranza di tali “documenti” è, difatti, di taglio estremamente generale, risolvendosi, quasi, seppur si tratti di norme interne di rango primario o sovranazionali giuridicamente vincolanti, in mere dichiarazioni di principi e linee guida, adottando, invero, un approccio ampio che punta a disciplinare l'intero insieme delle tecnologie riconducibili alla famiglia dell'intelligenza artificiale (è il caso, ad esempio, del Regolamento europeo) o, comunque, un loro settore esteso (si pensi alla, *supra* più volte citata, *Directive on automated decision-making* canadese).

⁶⁵ U. PAGALLO, *Algorithmic Decisions*, cit., p. 142; R. CALO, *Artificial Intelligence Policy*, cit., p. 430.

⁶⁶ L. FLORIDI, *The Ethics of Artificial Intelligence*, cit; W. BARFIELD, U. PAGALLO, *Advanced Introduction to Law and Artificial Intelligence*, Cheltenham, 2020.

⁶⁷ K. YEUNG, *The Regulation of Artificial Intelligence*, Londra, 2022; D.J. GUNKEL, *Robot Rules: Regulating Artificial Intelligence*, Londra, 2018.

Ciò è, a ben notare, 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. Siffatta circostanza ha sicuramente il pregio di fissare dei punti fermi – anche con norme tecniche rivolte ai pratici della disciplina – 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 – a mero titolo esemplificativo, per ciò che ivi rileva, basti pensare ai domini della medicina⁶⁸, della pubblica amministrazione⁶⁹, della giustizia⁷⁰ – dell'intelligenza artificiale che presentano peculiarità che non possono essere ignorate. Invero, talune applicazioni delle tecnologie intelligenti in contesti particolarmente delicati - come quelli appena menzionati - per il rango degli interessi coinvolti, necessiterebbero di una disciplina specifica, ritagliata sull'ambito di riferimento.

7. Riflessioni conclusive

Preme ribadire, nuovamente ed in conclusione, l'obiettivo fondamentale della presente ricerca: evidenziare come la rivoluzione connessa all'intelligenza artificiale chiami in causa la categoria giuridica dei diritti fondamentali, anche e soprattutto in termini di costruzione di nuovi diritti. 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 occuparsene⁷¹.

Si tratta di una percezione del tutto erronea: *in primis*, in quanto 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 proprie e personali competenze tecniche a riguardo; *in secundis*, in quanto diverse caratteristiche di tali tecnologie incidono direttamente sulla relazione tra uomo e macchine, tra uomo e natura. È il caso,

⁶⁸ DELOITTE, *The future of AI in healthcare. How AI will impact patients, clinicians and the pharmaceutical Industry*, 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*, 2021, pp. 1727-1776; T. DAVENPORT, R. KALAKOTA, *The potential for artificial intelligence in healthcare*, in *Future Healthcare Journal*, 2019, pp. 94-98.

⁶⁹ E. RAFFIOTTA, *L'erompre 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 (a cura di)*, 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*, 2022; A. PAJNO ET AL., *AI: profili giuridici. Intelligenza artificiale: criticità emergenti e nuove sfide per i giuristi*, in *Biolaw Journal*, 2019, p. 205; C. CASONATO, *Algoritmi e pubblica amministrazione: decisioni automatizzate e diritti*, in *Giornale di Diritto Amministrativo*, 2020, p. 545; G. SARTOR, *L'amministrazione algoritmica*, cit., p. 322.

⁷⁰ M. LUCIANI, *La decisione giudiziaria robotica*, in *Rivista AIC*, 3, 2018; T. SOURDIN, *Judge v Robot?: Artificial intelligence and judicial decision-making*, in *The University of New South Wales Law Journal*, 2018, pp. 1114–1133; F. FAGAN, S. LEVMORE, *The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion*, in *California Law Review*, 2019, pp. 1-37.

⁷¹ Sul punto, 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*, 2019, p. 195 ss.; S. QUINTARELLI (a cura di), *Intelligenza artificiale. Cos'è davvero, come funziona, che effetti avrà*, Torino, 2020.

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 risultati⁷².

A questo punto, è evidente come siffatte componenti tecniche portino ad un ridisegno del rapporto tra essere umano e realtà che lo circonda, da un lato, mettendo in dubbio la sua natura di agente esclusivo in seno a determinate attività in precedenza non automatizzabili, dall’altro, erodendo le possibilità di controllo *in toto* su prodotti di sua stessa creazione, elemento indiscusso prima dell’avvento delle nuove tecnologie.

Ergo, il rischio che la complessità tecnica insita nel funzionamento di tali sistemi nasconde la pregnanza della portata delle questioni intrinsecamente coinvolte è presente anche nella regolazione normativa del fenomeno. I primi tentativi a tal uopo, difatti, come si è visto nel corso del lavoro, si configurano per lo più quali norme prettamente tecniche, volte ad orientare l’azione degli specialisti che si occupano della progettazione e dello sviluppo delle applicazioni tecnologiche in esame.

La caratteristica di regole di tal fatta – 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 diritto⁷³. Enunciazioni di principio e richiami ad altri documenti – dal certo valore giuridico, *i.e.* concernenti aspetti valoriali essenziali *stricto sensu* – sono spesso del tutto assenti, o comunque meno frequenti che in testi normativi d’altro tipo.

Ciò potrebbe portare ad oscurare la circostanza che norme di tal genere - nel settore dell’intelligenza artificiale per ciò che ivi rileva – possano costituire strumento centrale per l’attuazione dei diritti fondamentali e dei principi generali dell’ordinamento. Le ragioni sono intuibili: si pensi, *ad exemplum*, 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 in principi generali nelle carte fondamentali, o dell’elaborazione dottrinale in materia⁷⁴.

Per quanto concerne l’intelligenza artificiale, alcuni dei testi normativi in materia dimostrano una certa consapevolezza di quanto appena affermato: è il caso, in particolare, del Regolamento europeo, che inquadra tra i suoi obiettivi fondamentali proprio la tutela dei diritti fondamentali.

Il presente lavoro compie un passo ulteriore e sostiene che quando i valori in gioco riguardino la protezione della sfera della personalità individuale nei confronti di un potere, *lato sensu* inteso – ovvero, pubblico o privato che sia, umano o tecnologico – l’elaborazione di norme seppur tecniche può rappresentare un indice circa la necessità di aggiornare, a livello almeno teorico, il catalogo dei diritti fondamentali con nuove tutele giuridiche di rango primario, a difesa di corrispondenti emergenti esigenze.

D’altro canto, solo raramente un nuovo diritto prende *ab origine* vita con l’enunciazione in costituzioni – e in tal caso previsioni di tal fatta non garantiscono di per sé una vera

⁷² 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, pp. 195–204; Z.C. LIPTON, *The mythos of model interpretability*, 2017.

⁷³ Ex multis, A. SANTOSUSSO, *Intelligenza artificiale, conoscenze neuroscientistiche e decisioni giuridiche*, in *Teoria e critica della regolazione sociale*, 2021, p. 189 ss.; A.G. VEDALDI, L. HANSEN, K.R. MÜLLER (a cura di), *Explainable AI: interpreting, explaining and visualizing deep learning*, Cham, 2019.

⁷⁴ Sul punto, ex plurimis, G.D. COMPORTI, *Codice dell’ambiente e responsabilità ambientale*, in *Rivista Quadrimestrale di Diritto dell’ambiente*, 2022, pp. 14-58; M. PIGNATTI, *La tutela dell’ambiente negli ordinamenti giuridici europei. Le valutazioni ambientali come strumenti di bilanciamento degli interessi e dei diritti*, in *DPCE online*, 2022, pp. 1319-1338.

effettività dello stesso *intra* ordinamento – comparendo, spesso, *in primis*, in disposizioni di dettaglio volte alla sua tutela in una specifica materia o in ambito giurisprudenziale, ove la positivizzazione solenne è solo eventuale e, nel caso, termine finale⁷⁵. Ciò è proprio il fenomeno che sta avvenendo in riferimento alla regolazione dell'intelligenza artificiale⁷⁶, relativamente agli ivi trattati tre nuovi diritti, il cui contenuto – più o meno compiuto – è stato discusso attraverso la disamina normativa *supra* effettuata e ove la possibilità di conflitto con prerogative altrui – questione di cui si accennava in apertura del lavoro – è aspetto naturale e certamente risolubile attraverso un bilanciamento tra i vari interessi rilevanti in concreto⁷⁷.

⁷⁵ 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*, 2019, p. 107 ss; P. GROSSI, *L'invenzione del diritto*, cit., p. 137; A. SIMONCINI, *Costituzionalismo e nuove tecnologie*, cit., p. 17.

⁷⁶ M. EBERS, S. NAVAS, *The EU Artificial Intelligence Act*, cit., p. 36; S. PENASA, *Intelligenza artificiale e diritti fondamentali*, cit., p. 28.

⁷⁷ Ad esempio, particolarmente delicata è la tensione tra spiegabilità e tutela del segreto commerciale; a tal uopo, la soluzione si colloca lungo la linea del bilanciamento, già sperimentata in ambito *antitrust*: il segreto industriale può essere preservato, ma non a costo del sacrificio del nucleo essenziale dei diritti dell'individuo, *in primis* il diritto di difesa. Invero, nella prassi, il ricorso a perizie tecniche eseguite sotto vincolo di riservatezza rappresenta, in tal senso, uno strumento di composizione equilibrata delle contrapposte esigenze (sul punto, U. PAGALLO, *Responsabilità e Intelligenza Artificiale*, Torino, 2021).