



Ordine internazionale e diritti umani

International Legal Order and Human Rights
Ordenamiento Jurídico Internacional y Derechos Humanos
Ordre Juridique International et Droits de l'Homme
Diretta da Claudio Zanghì, Lina Panella, Carlo Curti Gialdino

EDITORIALE
SCIENTIFICA

Focus tematico

15 dicembre 2025

Intelligenza artificiale e diritti umani: sfide e prospettive nel quadro giuridico europeo e internazionale

Dieci anni della Rivista "Ordine internazionale e diritti umani"

www.rivistaoidu.net

Rivista telematica - ISSN 2284-3531

Registrazione al Tribunale di Roma n. 46/2014 del 10 marzo 2014

ORDINE INTERNAZIONALE E DIRITTI UMANI

Focus tematico

Diretta da CLAUDIO ZANGHÌ, LINA PANELLA, CARLO CURTI GIALDINO

COMITATO SCIENTIFICO

Rafâa Ben Achour – Anthony Arnulf – Jean-Cristophe Barbato – Maria Caterina Baruffi – Andrea Biondi – Isabelle Bosse-Platière – Mouloud Boumghar – Laurence Burgogue-Larsen – Antonio Remiro Brotons – Ruggiero Cafari Panico – Giandonato Caggiano – Mario Pio Calogero – Andrea Cannone – Ida Caracciolo – Giovanni Cellamare – Consuelo Ramón Chornet – Gianluca Contaldi – Marcello Di Filippo – Zlata Drnas Clement – Emmanuel Decaux – Alejandro del Valle Gálvez – Angela Del Vecchio – Angela Di Stasi – Carlos Villan Duran – Osman El Hajjié – Juan Manuel Faramiñan – Gilbert – Pablo Antonio Fernandez Sanchez – Paolo Fois – Olivier de Frouville – Pietro Gargiulo – Giancarlo Guarino – Hajer Gueldich – Luis Miguel Hinojosa-Martínez – Paola Ivaldi – Slim Laghmani – Jaime Ferrer Lloret – Alessio Lo Giudice – Sergio Marchisio – Magdalena María Martín Martínez – Paolo Mengozzi – Claudia Morviducci – Bruno Nascimbene – Massimo Panebianco – Nicoletta Parisi – Piero Pennetta – Béatriz Pérez De La Heras – Stelios Perrakis – Yves Petit – Carlos Jimenez Piernas – Pietro Pustorino – Guido Raimondi – Fabián Salvioli – Tullio Scovazzi – Lucia Serena Rossi – Antonio Ruggeri – Francesco Seatzu – Attila Tanzi – Christian Tomuschat – Soledad Torrecuadrada García-Lozano – Sebastian Touzé – Takis Tridimas – Ennio Triggiani – Ugo Villani – Jorge E. Viñuales.

COMITATO DI REDAZIONE

Responsabile osservatori: Raffaele Cadin

Coordinatore di redazione: Francesco Battaglia

Componenti: Patrizia Accordino – Gabriele Asta – Maria Cristina Carta – Nicola Colacino – Giulia Colavecchio – Cristina Grieco – Luigino Manca – Gianfranco Gabriele Nucera – Antonio J. Palma – Francesca Perrini – Anna Pitrone – Luigi Prosperi – Valentina Zambrano.

Segreteria: Marco Bolognese – Pierfrancesco Breccia – Alberto Marchese – Luigi Zuccari.

Criteri di referaggio:

I contributi del focus sono stati sottoposti a referaggio in forma anonima.

Il presente *focus* tematico è stato pubblicato nell'ambito dei lavori della cattedra Jean Monnet *Digital Rights Empowerment Across Market in the European Union* (DREAM-EU), finanziata dall'Unione europea. Le opinioni e i pareri espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione Europea. Né l'Unione europea né l'autorità che concede la sovvenzione possono essere ritenute responsabili per essi.



**Co-funded by
the European Union**

Intelligenza artificiale e diritti umani: sfide e prospettive nel quadro giuridico europeo e internazionale

Dieci anni della Rivista “Ordine internazionale e diritti umani”

© Copyright - *Ordine internazionale e diritti umani*
Focus tematico inserito nel n. 5/2025
15 dicembre 2025

ISSN 2284-3531



Ordine internazionale e diritti umani

International Legal Order and Human Rights
Ordenamiento Jurídico Internacional y Derechos Humanos
Ordre Juridique International et Droits de l'Homme
Diretta da Claudio Zanghì, Lina Panella, Carlo Curti Gialdino

EDITORIALE
SCIENTIFICA

<http://www.rivistaoidu.net>

SOMMARIO FOCUS TEMATICO DEL N. 5/2025, 15 DICEMBRE 2025

INTELLIGENZA ARTIFICIALE E DIRITTI UMANI: SFIDE E PROSPETTIVE NEL QUADRO GIURIDICO EUROPEO E INTERNAZIONALE

Dieci anni della Rivista "Ordine internazionale e diritti umani"

- CLAUDIO ZANGHÌ, *I dieci anni della rivista "Ordine internazionale e diritti umani"*, pp. 1133-1135.
- RAFFAELE CADIN, *Presentazione degli Osservatori della Rivista "Ordine Internazionale e Diritti Umani"*, pp. 1136-1138.
- FRANCESCO BATTAGLIA, *La Cattedra Jean Monnet "Digital Rights Empowerment Across Market in the European Union (DREAM-EU)"*, pp. 1139-1141.
- GIANLUCA CONTALDI, *I profili generali dell'AI Act*, pp. 1142-1164.
- FIAMMETTA BORGIA, *Il modello europeo di regolazione dell'intelligenza artificiale: tra ambizione etica e ipertrofia normativa*, pp. 1165-1184.
- ROBERTO AMAGLIANI, *Intelligenza artificiale ed interesse superiore del minore*, pp. 1185-1194.
- FRANCESCO CIRAIOLO, *Credit scoring algoritmico e tutela della persona. Riflessioni intorno ai casi Schufa (CGUE, C-634/21) e D&B (CGUE, C-203/22)*, pp. 1195-1221.
- CRISTINA GRIECO, *Dal microtargeting politico ai deepfake: come le nuove tecnologie stanno riscrivendo la democrazia in europa*, pp. 1222-1234.
- LORENZO DI ANSELMO, *Intelligenza artificiale, lotta alla disinformazione online e tutela della democrazia nell'Unione europea alla luce dell'AI Act*, pp. 1235-1252.

- ALESSANDRO ROSANÒ, *Tu chiamali, se vuoi, sistemi di riconoscimento delle emozioni: considerazioni sulla disciplina posta in materia dal regolamento (UE) 2024/1689*, pp. 1253-1263.
- DARIA PERRONE, *Sistemi informatici esperti (SIE) e gestione del rischio nell'accertamento della responsabilità penale*, pp. 1264-1273.
- ANDREA CAPUTO, *Giudice-macchina e dignità umana: error 404*, pp. 1274-1285.
- MARCO SILIGATO, *Il settore della giustizia nell'AI Act: il dovere di trasparenza come "anidride solforosa"?*, pp. 1286-1297.
- ARIANNA BAVASTRELLI, *Intelligenza Artificiale e giustizia: alcune prospettive in ambito UE*, pp. 1298-1310.
- LUIGI PIGNA, *LA generativa e trattamento dei dati personali. La tutela dei diritti fondamentali dell'utente alla luce della recente iniziativa di meta per l'addestramento dell'LA*, pp. 1311-1328.
- DARIA LUISA PETRUCCI, *La tutela del diritto dell'individuo a un ambiente salubre nell'era dell'LA: sfide e prospettive*, pp. 1329-1342.
- ANDREA CASTALDO, *Intelligenza artificiale e diritti fondamentali. l'approccio "antropocentrico" tra Consiglio d'Europa e AI Act*, pp. 1343-1353.
- FABIANA MAGNOLO, *La costruzione di nuovi diritti fondamentali dinanzi alle sfide dell'intelligenza artificiale in un'ottica di (ri)valorizzazione del ruolo umano*, pp. 1354-1369.
- ISABELLA SALSANO, *Artificial Intelligence and the Protection of Cultural Heritage: Human Rights and the Role of International Law in Addressing Future Ethical Dilemmas*, pp. 1370-1386.
- ALESSIA DEL PIZZO, *Shaping fair ai: overcoming algorithmic bias and fostering inclusion*, pp. 1387-1403.
- PAOLO DALLE CARBONARE, *Dovere tributario e interesse fiscale: la difficile sopravvivenza del diritto alla riservatezza nell'era dell'intelligenza artificiale*, pp. 1404-1414.



Ordine internazionale e diritti umani

International Legal Order and Human Rights
Ordenamiento Jurídico Internacional y Derechos Humanos
Ordre Juridique International et Droits de l'Homme
Diretta da Claudio Zanghì, Lina Panella, Carlo Curti Gialdino

EDITORIALE
SCIENTIFICA

CLAUDIO ZANGHÌ*

I DIECI ANNI DELLA RIVISTA “ORDINE INTERNAZIONALE E DIRITTI UMANI”

Il mondo accademico italiano e non solo è stato dominato, per molti anni, dai cosiddetti “Baroni” con le relative scuole. Il sistema è stato certamente presente nel secolo passato quando le cattedre, e mi riferisco ovviamente a quelle di diritto internazionale, erano assai poche (una ventina) e permettevano quindi il permanere e lo sviluppo di tale sistema.

Successivamente, con l'aumento, smisurato, se posso esprimermi, delle Università in Italia e delle relative cattedre che hanno smembrato in più discipline quelle che originariamente erano una sola, la situazione di fatto che ne è conseguita ha certamente contribuito a smantellare il predetto regime. A prescindere da ogni considerazione negativa sullo stesso, è indubbio che il meccanismo si è riprodotto creando un fenomeno certamente non positivo per lo sviluppo della ricerca scientifica, che è quello della paternità di diverse riviste che, per ragioni di scuola di appartenenza universitaria e simili, finivano col pubblicare i contributi degli studiosi riconducibili alla medesima paternità escludendo, di fatto, qualunque contributo esterno. Ciò ha penalizzato inevitabilmente i giovani studiosi che non sempre rientravano in una delle suddette appartenenze e che, pertanto, hanno trovato concreta difficoltà nel pubblicare le loro ricerche, pur essendo noto, d'altra parte, che nell'attuale sistema universitario la valutazione dei giovani ricercatori è ormai direttamente connessa agli studi pubblicati nelle riviste ed alla qualità stessa di quest'ultime.

Ben consapevole della situazione ed appartenendo ad un'area (centro sud) dell'Italia nella quale era ben difficile per i giovani ricercatori trovare ospitalità in riviste del settore, ed avendo avuto la fortuna di entrare in relazione con molti colleghi, non solo italiani ma anche e prevalentemente stranieri, mi sono chiesto se fosse stato possibile escogitare una soluzione per dare spazio ai giovani ricercatori validi senza alcuna distinzione di appartenenza a scuole e ad aree.

Ero ben consapevole che i pesanti costi necessari per la stampa e la diffusione delle riviste sarebbero stati considerevolmente ridotti utilizzando la soluzione della rivista *online* ormai largamente diffusa nel settore scientifico e internazionale. Ciò era ed è particolarmente vero in funzione anche della tempistica che può essere realizzata mediante una rivista *online* rispetto alla tradizionale carta stampata. È noto, infatti, che i tempi per quest'ultima sono spesso assai lunghi e che di conseguenza uno studio elaborato da un autore attende spesso

* Professore emerito di Diritto internazionale, Università di Roma “La Sapienza”.

diversi mesi prima di essere pubblicato. La soluzione telematica permette, invece, di pubblicare i contributi in pochi giorni e quindi agevola enormemente, specie quando ciò è richiesto dal contenuto dello studio, dalla correlazione tra gli eventi internazionali avvenuti ed i commenti, da interpretazioni e valutazioni negli stessi scritti. Al contempo ero altrettanto consapevole che imbarcarsi nella istituzione di una rivista sarebbe stato un compito assai arduo perché è necessario che la rivista riceva i contributi da pubblicare, e che i suoi numeri siano pubblicati regolarmente non essendo certo un'operazione a termine bensì destinata a durare nel tempo. È noto, ad esempio, che già a quel tempo alcune riviste nascevano per morire poi dopo pochi mesi o anni. Potendo contare su un notevole numero di colleghi italiani e stranieri ho effettuato, quella che oggi si chiamerebbe una indagine di mercato, rivolgendomi a tanti colleghi (almeno 150) circa la loro volontà di contribuire all'impresa per capire fino a che punto l'idea avrebbe potuto avere un seguito con gli obiettivi sopra indicati.

La realizzazione di questa difficile impresa è stata possibile dopo aver acquisito i risultati ampiamente positivi con una lista, particolarmente estesa, di esplicita adesione all'iniziativa. Con l'aiuto prezioso della collega Lina Panella ci siamo lanciati in questa avventura concretatasi con la nascita della rivista *online* di cui oggi ricordiamo i 10 anni trascorsi. Alla rivista è stato attribuito il titolo di *Ordine Internazionale e Diritti Umani*. Il titolo, che potrà apparire strano per riunire in sé tematiche certamente vicine ma al contempo diversificate, trae origine dal titolo del Dottorato di ricerca, da anni esistente presso il Dipartimento di Scienze Politiche della Sapienza, Università di Roma, che ha frequentemente collaborato e collabora ancora oggi con le Università di Messina, Napoli, Salerno ed altre.

La duplicità dei temi che riuniscono l'ordinamento giuridico internazionale da una parte, e la tutela dei diritti umani dall'altra, consente alla Rivista di accogliere contributi che appartengono al mondo giuridico internazionale ed alle relazioni internazionali come pure a quello della tutela internazionale dei diritti dell'uomo. Tutti i contributi pubblicati nella Rivista sono sottoposti a preventivo referaggio che ne garantisce la rilevanza scientifica dal momento che la Rivista si avvale di numerosi esperti italiani e stranieri ai quali viene chiesto di esaminare il contributo in funzione della materia trattata.

Oltre ai contributi consueti, pubblicati in base agli autori in lingua italiana nonché in quelle europee di uso comune (francese e inglese e spagnolo), abbiamo deciso altresì di introdurre degli Osservatori il cui coordinamento è stato affidato e lo è ancora oggi, al collega Raffaele Cadin. È noto, infatti, che vi sono alcuni organismi internazionali che istituzionalmente si occupano della materia dei diritti umani e che pertanto meritano un aggiornamento periodico con relativo commento delle attività svolte. Con questo obiettivo si è così deciso di pubblicare ad ogni numero della Rivista uno o più specifici contributi dedicati alle attività svolte da ciascun organismo. Questi ultimi si sono poi accresciuti nel tempo.

La rivista, com'è consueto, è completata da un Comitato scientifico. Si è voluto evitare un pletorico comitato, che normalmente rimane soltanto un'indicazione nelle riviste conosciute, e si è preferito selezionare un numero ristretto di personalità scientifiche del mondo accademico italiano e straniero che hanno direttamente manifestato il loro interesse a seguire lo sviluppo della Rivista.

Il 15 Marzo del 2014 è stato così pubblicato *online* il primo numero della Rivista che si è assegnata una cadenza bimestrale con una pausa estiva, sempre alla stessa data del 15 dei mesi di marzo, maggio, giugno, ottobre e dicembre di ciascun anno. Negli anni successivi, alla condirezione con Lina Panella, si è aggiunto Carlo Curti Gialdino e così oggi rappresentiamo la triade che guida la Rivista assistita nel frattempo da un comitato di

redazione sotto la guida del collega Francesco Battaglia, composto da tanti giovani volenterosi, inseriti nell'ambito universitario, che si occupano di esaminare e seguire fino alla pubblicazione ogni contributo pubblicato.

Nel corso degli anni la Rivista ha accresciuto il suo ruolo di divulgazione della materia attraverso dei numeri "speciali" che hanno spesso riunito le relazioni di un intero convegno o di una ricerca anche pluridisciplinare. Ormai è divenuta una prassi consolidata ed almeno due numeri speciali l'anno vengono pubblicati.

Nel quadro del suo obiettivo primario di offrire a giovani ricercatori facilità di pubblicazione, senza incorrere in discriminazioni di appartenenza a scuole o Università, la Rivista ha dato vita ad una serie di "Quaderni", sempre pubblicati dopo il controllo scientifico dello scritto, e ad una collana di "Monografie", questa volta cartacee, che contribuiscono alla crescita dei giovani ricercatori. Ad oggi ben 6 "Quaderni" e 14 "Monografie" sono stati già pubblicati.

A 10 anni di distanza possiamo dire che la cadenza sopra richiamata è stata sempre rispettata dal primo numero fino a quello del 15 dicembre scorso. Ma oltre alla regolarità dei numeri nel corso del tempo mi piace sottolineare, quale merito di tutta la squadra che segue la Rivista, la ricchezza dei contenuti della stessa. Non sto certo ad elencare i tanti illustri personaggi che hanno contribuito al successo ma non posso ignorare che la quantità delle pagine per ogni numero della Rivista, in funzione anche del numero dei contributi pubblicati e dei commenti sui diversi Osservatori, è costantemente cresciuta raggiungendo centinaia di pagine, inimmaginabili in una rivista cartacea, e che solo in una rivista *online* sono consentite. È altrettanto un compiacimento per tutti constatare che dopo il primo triennio la Rivista, valutata dalla Commissione nazionale dell'ANVUR, ha ottenuto la classificazione nella categoria "A". Di conseguenza sono anche cresciute le richieste di pubblicazioni e la rilevanza della Rivista in Italia e dall'estero.

Non sta a me, specie in questa sede, far rimarcare il successo dell'iniziativa e la rilevanza assunta per continuità e contenuti. Il merito di tutto ciò è dovuto ai colleghi, ed ai giovani ricercatori e dottori di ricerca, che costantemente si prodigano per mantenere il ritmo dei numeri ed il livello delle pubblicazioni. È certo che senza l'aiuto di tutti l'iniziativa sarebbe stata destinata a fallire. Fortunatamente oggi non è così e possiamo quindi con gioia e soddisfazione di tutti celebrare i 10 anni di attività augurando lunga vita alla Rivista.



RAFFAELE CADIN*

PRESENTAZIONE DEGLI OSSERVATORI DELLA RIVISTA “ORDINE INTERNAZIONALE E DIRITTI UMANI”

Un caro saluto a tutte/i e un grazie ai Direttori della Rivista che mi hanno invitato a presentare gli Osservatori in questa importante occasione, in particolare alla Prof.ssa Lina Panella che è anche la premurosa padrona di casa.

Parlare dopo la magistrale presentazione della Rivista del Prof. Claudio Zanghì è nello stesso tempo facile e difficile. È certamente complicato non sfigurare davanti alla chiarezza esemplare dei suoi ragionamenti che come sempre rivelano una sensibilità nello stesso tempo profonda e sorprendente, ma d'altra parte è più semplice intervenire quando il senso e il significato della nascita della Rivista sono stati già mirabilmente tratteggiati nell'intervento di apertura del presente incontro.

E che non siano parole di circostanza lo dimostra il fatto che molte delle considerazioni appena fatte per la Rivista valgono anche per gli Osservatori, in particolare il tentativo, che oggi dopo dieci anni dalla nascita della Rivista può essere considerato riuscito, di rappresentare una sede editoriale autorevole aperta alla contribuzione scientifica di giovani studiose/i senza alcuna preclusione quanto alla scuola accademica di provenienza e/o di appartenenza.

A questo punto è però opportuno fare un passo indietro per (cercare di) spiegare cosa sono gli Osservatori della Rivista OIDU e perché non si tratta dei soliti contributi brevi di aggiornamento giuridico presenti in tante riviste, ma qualcosa di più e di diverso considerando anche che stiamo parlando di un Rivista *online* ad accesso libero.

In estrema sintesi vanno evidenziati i seguenti elementi di novità:

a) anche grazie al loro posizionamento grafico nella colonna nel lato destro della *home page* della Rivista e alla possibilità di aprire una pagina a loro dedicata cliccando sul logo di ciascuno di essi, gli Osservatori costituiscono una specie di rivista nella rivista che raccoglie tutti i contributi pubblicati nel corso degli anni;

b) l'identità di ogni Osservatorio è poi valorizzata dal fatto, naturalmente non occasionale, che i rispettivi curatori sono tra i principali esperti italiani e stranieri nell'ambito

* Professore Ordinario di Diritto Internazionale, Sapienza Università di Roma.

scientifico di riferimento, che sia la Corte Internazionale di Giustizia o una delle differenti corti regionali sui diritti umani per fare solo qualche esempio;

c) infatti, quasi tutti gli Osservatori, in particolari quelli presenti fin dalla nascita della Rivista, hanno carattere istituzionale e anche questo ne costituisce un valore aggiunto perché consente ai lettori della Rivista di essere costantemente aggiornati sulla pratica e sulla giurisprudenza delle principali entità, aventi natura giurisdizionale e non, competenti in materia di diritti umani *et similia*;

d) questo peraltro non significa che l'elenco degli Osservatori non sia stato incrementato in questo decennio in quanto, sia pure con molta cautela, nel corso degli anni ne sono stati introdotti di nuovi come quello sulle Organizzazioni non governative, ma anche a carattere tematico come quelli sul commercio internazionali e sulle nuove tecnologie.

La Rivista si ritrova quindi questo vero e proprio patrimonio di ben tredici Osservatori "specializzati", ognuno di essi con una propria storia e identità, ma che non avrebbero senso e rilevanza se non collegati al corpo editoriale centrale come gli Istituti specializzati lo sono all'ONU dando vita alla "famiglia" delle Nazioni Unite. Né va sottovalutata la dimensione umana degli Osservatori, i cui curatori, pur appartenendo a diverse generazioni, a differenti Paesi e anche a carriere non solo accademiche, costituiscono nel loro insieme una comunità di autorevoli studiosi accomunati dal fatto di essere cultori del diritto internazionale dei diritti umani.

Mi sia consentito qui di ricordare con grande affetto il Prof. Sergio García Ramírez, Presidente Emerito della Corte interamericana dei diritti umani e prolifico curatore per diversi anni del relativo Osservatorio, venuto a mancare nel gennaio 2024 al suo ritorno a Città del Messico dopo un viaggio a Roma per il conferimento del *Sapienza Human Rights Award*. I suoi puntuali contributi tematici sulla giurisprudenza della Corte di San José fanno sicuramente parte della storia editoriale non soltanto degli Osservatori, ma anche della nostra Rivista.

Una parola va anche spesa per sottolineare la qualità scientifica dei lavori pubblicati soprattutto, ma non esclusivamente, da giovani dottorandi, assegnisti e ricercatori, in una parola dalla giovane generazione delle internazionaliste e degli internazionalisti che spesso hanno vissuto la loro "prima volta" accademica pubblicando proprio nelle pagine degli Osservatori. Ad esempio, scorrendo a ritroso i contributi editi negli ultimi anni nell'Osservatorio da me curato sul Consiglio di sicurezza ho il piacere e l'emozione di ritrovare le prime prove scientifiche di tanti giovani studiosi che ora sono diventati stimati colleghi che insegnano in prestigiose università italiane e straniere.

Questa è poi la dimostrazione più sicura dell'entusiasmo e della cura che ha sempre caratterizzato prima l'elaborazione dei contributi da parte degli autori, con la supervisione dei curatori, e poi la loro pubblicazione quasi sempre notturna ad opera del *tandem* formato dallo scrivente e dell'insostituibile collega ed amico Prof. Francesco Battaglia, per tutti "Ciccio". Fermo restando che l'ultima parola è sempre spettata al Prof. Carlo Curti Gialdino nella sua qualità di co-Direttore della Rivista, nonché curatore dell'Osservatorio sulla Corte di giustizia dell'Unione Europea, ma soprattutto guida impareggiabile per tutti noi.

Quante storie più o meno curiose, quanti ricordi belli e purtroppo passati, quante nottate inebrianti di lavoro mi tornano in mente in occasione di questo decennale della Rivista! Ma lo sguardo deve essere rivolto al futuro non solo perché sotto il profilo editorial-scientifico il meglio deve sempre e per definizione ancora venire, ma soprattutto in ragione della "crisi d'umanità" che da diversi anni colpisce la comunità internazionale e il suo diritto. E, se è vero che i diritti umani hanno una capacità intrinseca di resistenza persino sotto le

rovine di Gaza, non si può negare che il ruolo promozionale di Riviste come “Ordine internazionale e Diritti Umani” rimane fondamentale e in prospettiva futura lo sarà sempre di più.



Ordine internazionale e diritti umani

International Legal Order and Human Rights
Ordenamiento Jurídico Internacional y Derechos Humanos
Ordre Juridique International et Droits de l'Homme
Diretta da Claudio Zanghì, Lina Panella, Carlo Curti Gialdino

EDITORIALE
SCIENTIFICA

FRANCESCO BATTAGLIA*

LA CATTEDRA JEAN MONNET “DIGITAL RIGHTS EMPOWERMENT ACROSS MARKET IN THE EUROPEAN UNION (DREAM-EU)”

La cattedra Jean Monnet 2024/2027 *Digital Rights Empowerment Across Market in the European Union (DREAM-EU)* presso il Dipartimento di Scienze Politiche dell'Università di Roma “La Sapienza” ha l'obiettivo di approfondire le tematiche collegate al mercato unico digitale, soprattutto dalla prospettiva della tutela dei diritti fondamentali. Un simile approccio è per certo in linea con la visione antropocentrica dell'azione dell'Unione in questo settore.

Invero, a partire dalla Strategia per il mercato unico digitale del 2015, l'Unione europea ha avviato un'intensa attività normativa volta a regolare i mercati digitali in un'ottica *human-centered*, al fine di trovare un giusto equilibrio tra il desiderio, da un lato, di incoraggiare l'innovazione tecnologica e la necessità, dall'altro lato, di garantire un elevato livello di protezione dei diritti fondamentali.

Gli strumenti chiave di tale azione normativa sono molteplici. Mi limito a ricordare il *Digital Services Act* (DSA), il *Digital Market Act* (DMA), il regolamento sull'intelligenza artificiale (AI Act), il *Data Act*, il *Data Governance Act*, le direttive sui contratti digitali, il regolamento europeo sulla libertà dei *media* e la direttiva sul diritto d'autore e il mercato digitale. Queste misure hanno delineato un quadro giuridico complesso, talvolta ritenuto fin troppo articolato. Ciò ha anche spinto, di recente, la Commissione europea a promuovere iniziative di semplificazione, come la proposta di adottare dei “*digital omnibus*”, i quali, tuttavia, potrebbero incidere negativamente sulla tutela dei diritti fondamentali.

A dieci anni dalla Strategia per il mercato unico digitale, quindi, è opportuno riflettere su come la posizione dei singoli sia cambiata in seguito alle numerose norme adottate dall'Unione per regolamentare i mercati digitali, anche alla luce delle loro prime applicazioni. Ciò non solo al fine di mettere in luce le principali novità introdotte, ma anche, anzi soprattutto, di identificare gli aspetti critici su cui appare necessario intervenire. Al momento, infatti, l'impressione di chi scrive è che, da un lato, il legislatore dell'Unione abbia compiuto uno sforzo significativo per rafforzare i diritti delle persone e per valorizzare la capacità di autodeterminazione degli utenti e dei consumatori; dall'altro lato, le grandi imprese del settore digitale stiano opponendo una notevole resistenza a queste nuove regole, sia per il

* Professore associato di Diritto dell'Unione europea, Università di Roma “La Sapienza”. Jean Monnet Chair 2024/2027 “Digital Rights Empowerment Across Market in the European Union (DREAM-EU)”.

timore che esse possano rimettere in discussione il modello di *business* su cui hanno costruito e consolidato la propria posizione di dominanza nei mercati di riferimento, sia perché le considerano eccessivamente complesse. Questa tensione, come detto, sta anche spingendo l'Unione europea a rivedere il proprio approccio in questo settore, adottando iniziative orientate alla semplificazione, come la recente proposta di *Digital Package*.

In tale contesto, la cattedra Jean Monnet DREAM-EU prevede per il triennio 2024-2027 una serie di attività di diversa natura sulle tematiche richiamate: attività didattiche, di ricerca e divulgative.

L'attività didattica si inserisce principalmente nell'offerta curriculare ed extracurriculare del Dipartimento di Scienze Politiche dell'Università di Roma "La Sapienza". Desidero, pertanto, esprimere un sentito ringraziamento alla Direttrice del Dipartimento di Scienze politiche, Prof.ssa Maria Cristina Marchetti, al Preside della Facoltà di Scienze politiche, Sociologia, Comunicazione, Prof. Pierpaolo D'Urso; ai presidenti delle Aree didattiche, Prof.ssa Paola Bozzao e Prof. Alessandro Guerra, alla coordinatrice del dottorato di ricerca in "Diritto pubblico, comparato e internazionale", Prof.ssa Giulia Caravale, e al suo predecessore, Prof. Fabio Giglioni; al coordinatore del dottorato in "Studi politici", Prof. Luca Micheletta. Un sentito ringraziamento, infine, desidero esprimerlo nei confronti di tutti i colleghi di Dipartimento, con i quali il confronto quotidiano è uno stimolo costante sul piano scientifico e umano.

In particolare, nell'ambito delle attività didattiche del Dipartimento, l'intero corso di *Diritto e politiche dell'Unione europea*, tenuto dal titolare della cattedra, è dedicato per tutto il triennio ai temi del mercato unico digitale e della tutela degli utenti. Sono, inoltre, previsti approfondimenti di 8 ore annue, sempre su tematiche collegate al progetto, all'interno dei corsi magistrali di *Governance sanitaria internazionale*, tenuto dalla Prof.ssa Valentina Zambrano, e di *Giustizia ambientale e lotta al cambiamento climatico*, impartito dal Prof. Gianfranco Gabriele Nucera. Un ringraziamento sincero va ad entrambi, non solo per aver accettato di figurare quali componenti della cattedra, ma anche per il loro ruolo di co-ideatori del progetto DREAM-EU nonché per l'infaticabile impegno profuso e il prezioso contributo fornito, sia nella fase progettuale sia in quella realizzativa.

La didattica sarà completata da cicli di seminari, organizzati annualmente su temi specifici, rivolti a studenti universitari, studenti delle scuole secondarie di secondo grado, dottorandi e professionisti.

Le attività di ricerca, invece, prevedono, tra le altre cose, una stretta collaborazione con la rivista *Ordine internazionale e diritti umani*, che in questa occasione celebra i suoi primi dieci anni di vita. A tal riguardo, colgo l'occasione per ringraziare i co-direttori Claudio Zanghi, Carmela Panella e Carlo Curti Gialdino, ai quali mi lega un profondo affetto. Tra le iniziative di ricerca già avviate, si ricorda, ad esempio, l'istituzione all'interno della *Rivista* di un osservatorio *Nuove tecnologie e diritti fondamentali*, realizzato con il prezioso sostegno del collega e amico Raffaele Cadin, che sin dalla nascita di *Ordine internazionale e diritti umani* dirige magistralmente la sezione dedicata agli osservatori.

Infine, nell'ambito della cattedra DREAM-EU sono state promosse diverse iniziative divulgative, tra cui un *blog* dedicato alle tematiche del progetto, pensato come spazio per approfondire questioni di stretta attualità, sia sul piano giurisprudenziale sia su quello normativo, stimolando riflessioni e dibattiti tra studiosi e cittadini interessati. Nel *blog* sono attivamente coinvolti anche gli studenti dei corsi di laurea interessati.

Per concludere, ma non per ordine di importanza, desidero ringraziare tutto il personale tecnico e amministrativo del Dipartimento di Scienze politiche della Sapienza per

l'instancabile e paziente supporto, nonché il dott. Lorenzo Di Anselmo ed il dott. Luigi Pigna, rispettivamente borsista post-doc in Diritto dell'Unione europea e dottorando di ricerca in Diritto pubblico, comparato e internazionale, presso il Dipartimento di Scienze politiche, entrambi giovani e promettenti studiosi.

Per ulteriori informazioni sulle iniziative promosse dalla cattedra, si fa rinvio al sito ufficiale www.dreamen.net, progettato con il contributo volontario e “spontaneo” di Antonio Battaglia.



GIANLUCA CONTALDI*

I PROFILI GENERALI DELL'AI ACT

SOMMARIO: 1. Il percorso per l'adozione del regolamento europeo sull'intelligenza artificiale. – 2. La base giuridica del regolamento. – 3. La definizione di intelligenza artificiale e l'ambito di applicazione della normativa europea. – 4. La piramide del rischio. – 5. Il rischio inaccettabile. – 6. I sistemi ad alto rischio. – 7. La *governance* dell'IA. – 8. Un approccio timido del legislatore europeo?

1. *Il percorso per l'adozione del regolamento europeo sull'intelligenza artificiale*

Anche se la proposta di adozione di un regolamento elaborata dalla Commissione europea è stata presentata nel 2021¹, l'intenzione delle istituzioni europee di occuparsi di intelligenza artificiale è più risalente. Fin dalla comunicazione posta a base del c.d. mercato unico digitale, in effetti, la Commissione parla di «un'economia industriale e di servizio intelligente» e constata che «(l)a digitalizzazione offre possibilità senza precedenti anche a altri comparti dell'economia, come i trasporti (ad es., sistemi di trasporto intelligenti) o l'energia (ad es., reti intelligenti, contatori intelligenti)»².

Ovviamente, siamo nel 2015 e, con l'appellativo di “intelligente”, l'esecutivo europeo intende riferirsi solo ai servizi e alle reti concepite per sfruttare al meglio le potenzialità di un dispositivo ovvero di organizzare in maniera ottimale il carico di una rete energetica. A conferma di questo assunto, è agevole osservare che, nel testo inglese, il termine impiegato è, – banalmente – «smart»; e non quello di *intelligent*, abitualmente impiegato per riferirsi ai processi articolati di elaborazioni dati e produzione di output.

Di intelligenza artificiale, in senso stretto, si inizia a parlare solo nella revisione della Comunicazione adottata due anni prima. Nel corso del 2017, l'esecutivo europeo approccia,

* Professore ordinario di Diritto dell'Unione europea, Università degli Studi di Macerata.

¹ Proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione {SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}, Bruxelles, 21.4.2021 COM(2021) 206 final 2021/0106 (COD).

² Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, Strategia per il mercato unico digitale in Europa, {SWD(2015) 100 final}, Bruxelles, 6.5.2015, COM(2015) 192 final, p. 15.

per la prima volta, il tema della responsabilità civile discendente dall'uso di sistemi di intelligenza artificiale, avvicinandola alla disciplina – all'epoca già esistente³ – della responsabilità per danno da prodotto⁴.

Successivamente, nel mese di aprile del 2018, la Commissione ha iniziato ad accompagnare le trattazioni concernenti la responsabilità dell'uso dei sistemi di intelligenza artificiale con la protezione dei diritti umani e con la necessità di porre l'uomo al centro dei progressi tecnologici nel settore digitale. Nella comunicazione «Intelligenza artificiale per l'Europa», la Commissione si propone, infatti, di rafforzare la leadership dell'Unione europea nell'intelligenza artificiale. A tal fine, la Commissione prevede di stanziare i fondi necessari per prepararsi ai cambiamenti socio-economici che la diffusione di tale tecnologia comporta⁵. E ancor più chiaramente, nel corso del mese di dicembre 2018, afferma che lo sfruttamento delle tecnologie dell'intelligenza artificiale dovrà essere accompagnato da principi etici da elaborare⁶. Pertanto, già tre anni prima di formulare la proposta di regolamento l'organo esecutivo ha ben presente che il futuro atto normativo dovrà attuare un bilanciamento tra i principi della responsabilità oggettiva, tipica dei danni da prodotti, ed i diritti fondamentali⁷.

In seguito, il processo legislativo ha subito un'accelerazione. Questa è stata indubbiamente determinata da una serie di fattori concomitanti. Innanzitutto, dallo scoppio della pandemia da COVID, la quale ha provocato un massiccio ricorso alle nuove tecnologie in diversi settori, quale strumento sostitutivo degli incontri e dei contatti personali. Si è così assistito, forse anche oltre le necessità dettate dalla pandemia, a una forte spinta verso la digitalizzazione e a un massiccio ricorso al «lavoro agile». Va poi menzionata la grave crisi economica e l'invasione dell'Ucraina da parte della Russia, che ha implicato la necessità di

³ La prima disciplina della responsabilità per danno da prodotto difettoso era contenuta nella direttiva 85/374/CEE del Consiglio del 25 luglio 1985 relativa al ravvicinamento delle disposizioni legislative, regolamentari ed amministrative degli Stati Membri in materia di responsabilità per danno da prodotti difettosi, in GUUE L 210 del 7.8.1985. Questa è stata ora abrogata dalla direttiva (UE) 2024/2853 del Parlamento europeo e del Consiglio, del 23 ottobre 2024, sulla responsabilità per danno da prodotti difettosi, che abroga la direttiva 85/374/CEE del Consiglio, in GUUE L, 2024/2853, 18.11.2024, che è stata emanata proprio per adattare la disciplina originaria alle evoluzioni del mondo digitale.

⁴ Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, sulla revisione intermedia dell'attuazione della strategia per il mercato unico digitale. Un mercato unico digitale connesso per tutti, {SWD(2017) 155 final}, Bruxelles, 10.5.2017, COM(2017) 228 final, p. 13: «La Commissione esaminerà l'eventuale opportunità di adeguare l'attuale quadro giuridico per tenere conto dei nuovi sviluppi tecnologici (ad esempio nei settori della robotica, dell'intelligenza artificiale e della stampa 3D), in particolare dal punto di vista della responsabilità civile, alla luce dei risultati della valutazione in corso sulla direttiva relativa alla responsabilità per danno da prodotti difettosi e sulla direttiva “macchine”».

⁵ Comunicazione della Commissione, L'intelligenza artificiale per l'Europa, Bruxelles, 25.4.2018, COM(2018) 237 final: «Per gestire la trasformazione dell'IA, i lavoratori il cui impiego sta cambiando o può scomparire a causa dell'automazione devono avere tutte le opportunità di acquisire le competenze e le conoscenze necessarie per padroneggiare la nuova tecnologia e ricevere sostegno durante la transizione nel mercato del lavoro. Questo atteggiamento proattivo e incentrato sull'investire nelle persone è il fondamento di un approccio all'IA inclusivo e antropocentrico, che richiederà un investimento notevole».

⁶ Comunicazione della Commissione al Parlamento europeo, al Consiglio europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, Piano coordinato sull'intelligenza artificiale, Bruxelles, 7.12.2018, COM(2018) 795 final, p. 7.

⁷ V. M. BASSINI, *Oggetto, campo di applicazione e ambito territoriale*, in O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI (a cura di), *La disciplina dell'intelligenza artificiale*, Milano, 2025, p. 113 ss., spec. 114 ss.; M. ALMADA, N. PETIT, *The EU AI Act: between the Rock of Product Safety and the Hard Place of Fundamental Rights*, in *Common Market Law Review*, 2025, p. 85 ss., secondo i quali la regolamentazione dell'AI Act combina norme sulla sicurezza dei prodotti e tutela dei diritti fondamentali.

riconvertire le fonti di approvvigionamento. Pure la crisi climatica e la correlata necessità di ridurre l'impiego delle fonti fossili hanno contribuito a rendere necessaria una normativa sull'utilizzo delle nuove tecnologie.

In questo contesto, il ricorso all'intelligenza artificiale è stato sempre più considerato quale un volano per lo sviluppo delle economie moderne. La stessa Commissione, nel Libro bianco sull'intelligenza artificiale, ha riconosciuto che: «I progressi compiuti nell'ambito del calcolo e la crescente disponibilità di dati sono pertanto fattori determinanti per l'attuale crescita dell'IA. L'Europa può combinare i suoi punti di forza industriali e tecnologici con un'infrastruttura digitale di elevata qualità e un quadro normativo basato sui suoi valori fondamentali per diventare un leader mondiale nell'innovazione nell'economia dei dati e nelle sue applicazioni (...). Su questa base l'Europa può sviluppare un ecosistema di IA che consenta alla sua società e alla sua economia nel loro complesso di godere dei benefici apportati dalla tecnologia»⁸.

Tutti questi fattori hanno dunque fatto maturare la necessità e la volontà di creare un ambiente favorevole allo sviluppo dell'intelligenza artificiale nell'Unione europea.

L'AI ACT è, pertanto, scaturito da queste molteplici esigenze. In definitiva, per una normativa così articolata e complessa, il procedimento legislativo è stato anche rapido. Rispetto ad una proposta della Commissione presentata in data 21 aprile 2021, il testo definitivo è stato approvato nel giro di tre anni, dal momento che lo stesso è stato licenziato dal Parlamento europeo e dal Consiglio, rispettivamente, nei mesi di marzo e maggio 2024⁹.

L'AI Act è senza dubbio uno strumento legislativo di fondamentale importanza nella costruzione del mercato unico digitale¹⁰. Come ha affermato Mathieu Michel, all'epoca sottosegretario di Stato belga per la Digitalizzazione: «L'adozione del regolamento sull'IA rappresenta una pietra miliare per l'Unione europea. Questo atto legislativo storico (...) affronta una sfida tecnologica globale che crea anche opportunità per le nostre società ed economie. Con il regolamento sull'IA l'Europa pone l'accento sull'importanza della fiducia, della trasparenza e della responsabilità nell'utilizzo delle nuove tecnologie, insieme alla garanzia che questa tecnologia in rapida evoluzione possa prosperare e stimolare l'innovazione europea»¹¹. In effetti, in questa dichiarazione, non a torto riportata da vari

⁸ Libro bianco sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia, Bruxelles, 19.2.2020, COM(2020) 65 final, p. 2.

⁹ Più dettagliatamente: il 21 aprile 2021 la Commissione europea ha presentato la propria proposta (Proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale), cit.; il 14 giugno 2023, il Parlamento Europeo ha proposto gli emendamenti; il 2 febbraio 2024 si è passati attraverso l'approvazione del COREPER; il 13 marzo 2024 c'è stato il via libera da parte del Parlamento Europeo; il 21 maggio 2024, il Consiglio approva la versione finale del regolamento. Il 12 luglio il regolamento 2024/1689 è pubblicato nella Gazzetta Ufficiale dell'Unione europea: Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale), in GUUE L, 2024/1689, 12.7.2024. La nuova disciplina si applicherà a far data dal 2 agosto 2026 (per riferimenti più precisi, in relazione alle diverse parti del regolamento v. la nota 63).

¹⁰ Cfr. B. NASCIBENE, *Il mercato unico digitale è una nuova frontiera dell'integrazione europea?*, in *rivista.eurojus*, 3/2025, p. 1 ss., 10, nel decennio 2020-2030 individua tre atti come fondamentali per la costruzione del mercato unico digitale (il *Digital Markets Act*, il *Digital Services Act* e, appunto, l'*AI Act*).

¹¹ La dichiarazione riportata nel testo è desunta dal testo del comunicato stampa del Consiglio dell'Unione europea, al momento dell'approvazione da parte dell'istituzione del testo finale del regolamento (v. il sito: <https://www.consilium.europa.eu/it/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/>).

commentatori, viene sintetizzata, in poche parole, la volontà del legislatore europeo sia di adottare uno standard normativo globale, sia di creare un'opportunità di sviluppo economico e, al contempo, di dare vita ad un ambiente digitale sicuro ed affidabile.

Nonostante la centralità riconosciuta dell'AI Act occorre, tuttavia, chiedersi se sia sufficiente un atto normativo per conquistare un ruolo di primo piano nello sviluppo dell'intelligenza artificiale. In questo momento, infatti, si assiste ad una sorta di corsa alle tecnologie dell'intelligenza artificiale e ad un affannoso tentativo dei legislatori, nazionali ed internazionali, di fornire una regolamentazione che tenga conto non solo delle esigenze economiche, ma anche dell'impatto di questa tecnologia sul godimento dei diritti fondamentali.

2. La base giuridica del regolamento

Il regolamento si fonda su una duplice base giuridica¹². Questa è composta, rispettivamente, dall'articolo 114 TFUE, che come è noto riguarda «le misure relative al ravvicinamento delle disposizioni legislative, regolamentari ed amministrative degli Stati membri che hanno per oggetto l'instaurazione ed il funzionamento del mercato interno»; e dall'articolo 16 TFUE, che garantisce il diritto di ogni persona alla protezione dei dati di carattere personale.

La doppia base giuridica è sintomatica delle molteplici finalità perseguite dalle istituzioni europee: da un lato, lo sviluppo del mercato interno, teso a favorire una transizione digitale e ad affermare una sorta di sovranità europea sui dati; e, dall'altro, il rispetto dei diritti fondamentali ed in particolare quello alla tutela dei dati personali.

Tale duplice obiettivo emerge fin dal primo considerando, il quale prevede espressamente che: «Lo scopo del presente regolamento è migliorare il funzionamento del mercato interno istituendo un quadro giuridico uniforme [...] in conformità dei valori dell'Unione, promuovere la diffusione di un'intelligenza artificiale (IA) antropocentrica e affidabile, garantendo nel contempo un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea, compresi la democrazia, lo Stato di diritto e la protezione dell'ambiente, proteggere contro gli effetti nocivi dei sistemi di IA nell'Unione, nonché promuovere l'innovazione. Il presente regolamento garantisce la libera circolazione transfrontaliera di beni e servizi basati sull'IA, impedendo così agli Stati membri di imporre restrizioni allo sviluppo, alla commercializzazione e all'uso di sistemi di IA, salvo espressa autorizzazione del presente regolamento».

Il contemperamento tra i due obiettivi che abbiamo sopra indicato (sviluppo del mercato interno e tutela dei diritti fondamentali) dovrebbe essere garantito dalla ricerca di una “via europea all'intelligenza artificiale”, mediante la creazione di un ecosistema di fiducia. I principi di base del regolamento – e la stessa tassonomia del rischio – mirano evidentemente

¹² Sulla base giuridica e sugli obiettivi generali del regolamento v. C. SCHEPISI, *Le “dimensioni” della regolazione dell'intelligenza artificiale nella proposta di regolamento della Commissione*, in V. SALVATORE (a cura di), *Digitalizzazione, intelligenza artificiale e tutela della salute nell'Unione europea*, Torino, 2023, p. 29 ss.; M. CARTA, *Il Regolamento UE sull'Intelligenza Artificiale: alcune questioni aperte*, in *rivista.eurojus.it*, n. 3/2024, p. 188 ss., 192; G. FINOCCHIARO, *Diritto dell'intelligenza artificiale*, Bologna, 2024, p. 25 ss., 54 ss.

a rafforzare la fiducia dei cittadini nell'uso, che può essere garantita solo da un'assoluta trasparenza dei sistemi e da un costante controllo umano sul funzionamento degli stessi¹³.

Per raggiungere tali obiettivi, il regolamento si basa sui principi di proporzionalità, trasparenza e sicurezza. Per proporzionalità si intende un approccio rapportato al rischio dei singoli sistemi, in funzione agli usi per i quali sono concepiti, in modo da non creare restrizioni eccessive al commercio. La trasparenza indica invece che i sistemi di IA devono essere progettati e sviluppati in modo da garantire il percorso logico che ha determinato l'assunzione delle decisioni. Infine, la sicurezza è il principio per cui i sistemi di IA devono essere progettati e sviluppati in modo da ridurre i rischi per la sicurezza e la salute delle persone, per i diritti fondamentali e per gli interessi pubblici.

Pur condividendo tale molteplicità di scopi, viene tuttavia da chiedersi se la base giuridica impiegata (artt. 114 e 16 TFUE) non finisca per essere, in un certo senso, limitativa a fronte della pluralità degli obiettivi perseguiti. In effetti, pur rappresentando i dati personali la base per lo sviluppo dei sistemi di intelligenza artificiale, sembra tuttavia criticabile limitare a tale aspetto le esigenze di protezione dei diritti fondamentali. Altri e non meno importanti diritti potrebbero infatti risultare pregiudicati da uno sviluppo incontrollato dell'intelligenza artificiale, quali il diritto di espressione, il principio di democrazia, il divieto di discriminazioni (quali possono essere quelle discendenti dai *bias* dei dati utilizzati per alimentare il sistema di intelligenza artificiale), la protezione dei minori e quella dei consumatori¹⁴.

Non c'è, d'altro canto, alcun dubbio che il legislatore europeo fosse pienamente consapevole di questi rischi. Lo stesso Consiglio, nelle conclusioni della Presidenza nel corso del 2020 dirette ad incoraggiare la Commissione a presentare una proposta di atto legislativo, manifesta chiaramente l'esigenza che i sistemi di intelligenza artificiale siano rispettosi dei diritti fondamentali. Si può infatti leggere in dette conclusioni che: «Sottolineiamo che la progettazione, lo sviluppo, la diffusione e l'utilizzo dell'IA devono rispettare appieno i diritti fondamentali e le norme giuridiche vigenti. È opportuno applicare lo stesso livello di protezione al mondo digitale e a quello fisico. Sottolineiamo che, ai sensi dell'articolo 52, paragrafo 1, della Carta, eventuali limitazioni all'esercizio dei diritti e delle libertà riconosciuti dalla Carta possono essere apportate solo laddove siano necessarie e rispondano effettivamente a una finalità di interesse generale riconosciuta dall'UE o all'esigenza di tutelare i diritti e le libertà altrui, nel rispetto del principio di proporzionalità, e devono essere previste dalla legge e rispettare il contenuto essenziale dei diritti e delle libertà fondamentali. La normativa pertinente che dà attuazione ai diritti fondamentali potrebbe essere messa in discussione a causa della complessità e dell'opacità di talune applicazioni di IA, i cui risultati richiederebbero esperienza e processi specializzati per essere compresi e controllati»¹⁵.

E la medesima Commissione europea, nella proposta presentata nel mese di aprile 2021, precisava che: «l'utilizzo dell'IA con le sue caratteristiche specifiche (ad esempio

¹³ Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, *Creare fiducia nell'intelligenza artificiale antropocentrica*, COM(2019) 168 final, Bruxelles, 8.4.2019, spec. par. 2.2.

¹⁴ Cfr. F. DONATI, *Diritti fondamentali e algoritmi nella Proposta di Regolamento sull'intelligenza artificiale*, in A. PAJNO, F. DONATI, A. PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, Bologna, 2022, p. 111 ss.; A. ODDENINO, *Intelligenza artificiale e tutela dei diritti fondamentali: alcune nozioni critiche sulla recente proposta di regolamento della UE, con particolare riferimento all'approccio basato sul rischio e al pericolo di discriminazione algoritmica*, ivi, p. 165 ss.; C. GRIECO, *Intelligenza artificiale e tutela degli utenti nel diritto dell'Unione europea*, Napoli, 2023.

¹⁵ Consiglio dell'Unione europea, Conclusioni della presidenza - La Carta dei diritti fondamentali nel contesto dell'intelligenza artificiale e della trasformazione digitale, Bruxelles, 21 ottobre 2020, 11481/20, punti 10-11, p. 7 s.

opacità, complessità, dipendenza dai dati, comportamento autonomo) può incidere negativamente su una serie di diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea. La presente proposta mira ad assicurare un livello elevato di protezione di tali diritti fondamentali e ad affrontare varie fonti di rischio attraverso un approccio basato sul rischio chiaramente definito. Definendo una serie di requisiti per un'IA affidabile e di obblighi proporzionati per tutti i partecipanti alla catena del valore, la presente proposta migliorerà e promuoverà la protezione dei diritti tutelati dalla Carta: il diritto alla dignità umana (articolo 1), al rispetto della vita privata e alla protezione dei dati di carattere personale (articoli 7 e 8), alla non discriminazione (articolo 21) e alla parità tra donne e uomini (articolo 23). Essa mira a prevenire un effetto dissuasivo sui diritti alla libertà di espressione (articolo 11) e alla libertà di riunione (articolo 12), nonché ad assicurare la tutela del diritto a un ricorso effettivo e a un giudice imparziale, della presunzione di innocenza e dei diritti della difesa (articoli 47 e 48), così come il principio generale di buona amministrazione¹⁶.

Vista l'attenzione ai diritti fondamentali, vale la pena domandarsi se non sarebbero stati opportuni ulteriori normativi, per fornire una base giuridica specifica proprio alla parte relativa alla tutela di questi¹⁷.

Il punto è tuttavia che, pur reperendo altre disposizioni all'interno del Trattato istitutivo che fanno chiaramente riferimento ai valori fondanti dell'Unione e ai diritti fondamentali, queste non sono tuttavia attributive di competenza. Non è, infatti, attributivo di competenza l'art. 2 TUE, il quale si limita a stabilire quali siano i valori fondanti dell'Unione. Né può essere attributivo di competenza l'art. 6, TUE, il quale, pur inserendo nel contesto del diritto primario, sia la Carta dei diritti fondamentali e le disposizioni in questa contenute (art. 6, par. 1), sia i diritti fondamentali desumibili dalla Convenzione europea dei

¹⁶ Proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale), cit., p. 13.

¹⁷ Tra i diritti fondamentali che potrebbero venire lesi dall'impiego di sistemi di intelligenza artificiale, dovrebbe includersi anche la proprietà intellettuale, anche se questa non è affatto menzionata nel passo sopra riportato della proposta della Commissione europea. Il problema si pone soprattutto per l'IA generativa, la quale raccoglie in via autonoma immagini e testi dal web, per poi elaborarne di nuovi. Sotto questo aspetto, la tutela del diritto d'autore sovviene in una duplice prospettiva. Il primo aspetto concerne la tutela del diritto d'autore a fronte di testi ed immagini diffuse sul web, che potrebbero essere tutelati dalle norme sulla proprietà intellettuale. In secondo luogo, c'è il problema di capire in quale misura le opere create con sistemi di IA generativa possano, a propria volta, accedere alla tutela autoriale. Il regolamento, nella misura in cui disciplina i sistemi di intelligenza artificiale generativa (denominati modelli di IA per finalità generali), prende posizione sul primo aspetto; v., in particolare, il considerando 106: «A tal fine, i fornitori di modelli di IA per finalità generali dovrebbero mettere in atto una politica volta a rispettare il diritto dell'Unione in materia di diritto d'autore e diritti connessi, in particolare per individuare e rispettare la riserva dei diritti espresse dai titolari dei diritti a norma dell'articolo 4, paragrafo 3, della direttiva (UE) 2019/790. Qualsiasi fornitore che immetta sul mercato dell'Unione un modello di IA per finalità generali dovrebbe rispettare tale obbligo, indipendentemente dalla giurisdizione in cui hanno luogo gli atti pertinenti in materia di diritto d'autore alla base dell'addestramento di tali modelli di IA per finalità generali». Al contrario, il regolamento non si occupa per nulla del secondo tema ovvero della tutela del diritto di autore per le opere generate da sistemi di IA. In dottrina, cfr. G. FINOCCHIARO, *Diritto dell'intelligenza artificiale*, cit., p. 121 ss.; C. GRIECO, *Intelligenza artificiale e diritto d'autore: riflessioni a partire dalla recente prassi interna ed europea*, in *Studi sull'integrazione europea*, 2024, p. 239 ss.; A. AMIDEI, *La "creatività computazionale"? Quale tutela autoriale per le opere dell'ingegno AI-generated?*, in U. RUFFOLO, A. AMIDEI, *Diritto dell'intelligenza artificiale*, Vol. II, *Proprietà industriale e intellettuale. Corp Techb. Giustizia predittiva. Transumanesimo. AI generativa. Metaverso*, Roma, 2024, p. 79 ss.; ID., *AI generativa, libertà di manifestazione del pensiero e diritto d'autore*, ivi, p. 139 ss.

diritti dell'uomo e dalle tradizioni costituzionali comuni (art. 6, par. 3), è tuttavia esplicito nel precisare che dette fonti non sono attributive di competenza¹⁸.

Neppure possono essere ritenute disposizioni attributive di competenza l'art. 19 TUE, che si limita a descrivere il sistema giurisdizionale europeo, ma risulta, di per sé, privo di attribuzione di poteri normativi specifici alle istituzioni europee¹⁹ e l'art. 8 TFUE, che pure sancisce il principio di non discriminazione, da tenere indubbiamente presente nella conformazione degli algoritmi, senza tuttavia apprestare alcun procedimento legislativo per l'adozione di atti a tutela dello stesso²⁰.

Quindi, il riferimento alle disposizioni sopra menzionate, che trattano rispettivamente dell'armonizzazione necessaria per il funzionamento del mercato interno e della tutela dei dati personali, è chiaramente limitativo rispetto all'obiettivo prefissato dal legislatore europeo; ma lo stesso risulta chiaramente non superabile, data l'assenza di altre previsioni attributive di competenza all'interno del Trattato istitutivo.

3. La definizione di intelligenza artificiale e l'ambito di applicazione della normativa europea

Il regolamento adotta una nozione di intelligenza artificiale che, per certi versi, è più descrittiva dei caratteri dei sistemi che vi fanno ricorso, che realmente definitoria della tecnologia stessa.

¹⁸ Per la precisione, l'art. 6, par. 2, secondo comma, TUE, precisa che «Le disposizioni della Carta non estendono in alcun modo le competenze dell'Unione definite nei trattati». Viceversa, la circostanza che l'art. 6, par. 3, del Trattato non sia una disposizione attributiva di competenza è stata esplicitata dalla Corte di giustizia nel noto parere 2/94, *Adesione della Comunità alla Convenzione per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali*, ECLI:EU:C:1996:140, punti 32-34: «Si deve anzitutto ricordare che la rilevanza del rispetto dei diritti dell'uomo è stata sottolineata in varie dichiarazioni degli Stati membri e delle istituzioni comunitarie (richiamate al punto III.5 della parte in fatto). Un riferimento al rispetto di tali diritti è contenuto anche nel preambolo dell'Atto unico europeo, nonché nel preambolo e negli artt. F, n. 2, J.1, n. 2, quinto trattino, e K.2, n. 1, del Trattato sull'Unione europea. L'art. F precisa, inoltre, che l'Unione rispetta i diritti fondamentali quali sono garantiti, in particolare, dalla Convenzione. L'art. 130 U del Trattato prevede, al n. 2, che la politica della Comunità nel settore della cooperazione allo sviluppo contribuisce all'obiettivo del rispetto dei diritti dell'uomo e delle libertà fondamentali. Si deve rilevare inoltre che, secondo costante giurisprudenza, i diritti fondamentali fanno parte integrante dei principi generali del diritto dei quali la Corte garantisce l'osservanza. A tal fine la Corte si ispira alle tradizioni costituzionali comuni agli Stati membri e alle indicazioni fornite dai trattati internazionali relativi alla tutela dei diritti dell'uomo a cui gli Stati membri hanno cooperato e aderito. In tale contesto la Corte ha precisato che la Convenzione riveste un significato particolare. Se il rispetto dei diritti dell'uomo costituisce, quindi, un requisito di legittimità degli atti comunitari, si deve tuttavia rilevare che l'adesione alla Convenzione determinerebbe una modificazione sostanziale dell'attuale regime comunitario di tutela dei diritti dell'uomo, in quanto comporterebbe l'inserimento della Comunità in un sistema istituzionale internazionale distinto, nonché l'integrazione del complesso delle disposizioni della Convenzione nell'ordinamento giuridico comunitario».

¹⁹ Pur essendo indubbio il ruolo della Corte di giustizia nell'introduzione e nella tutela dei diritti fondamentali nell'ordinamento giuridico dell'Unione europea: cfr. F. CAPOTORTI, *Il diritto comunitario non scritto*, in *Dir. com. sc. int.*, 1983, p. 409 ss.; A. DAVI, *Comunità europee e sanzioni economiche internazionali*, Napoli, 1993, p. 332 ss.; C. CURTI GIALDINO, *Art. 19 TUE*, in C. CURTI GIALDINO (a cura di), *Codice dell'Unione europea operativo*, Roma, 2012, p. 226 ss., 236 s.

²⁰ Per questa riflessione sull'art. 8 TFUE cfr. A. ADINOLFI, *Processi decisionali automatizzati e diritto antidiscriminatorio dell'Unione europea*, in A. ADINOLFI, A. SIMONCINI (a cura di), *Protezione dei dati personali e nuove tecnologie. Ricerca interdisciplinare sulle tecniche di profilazione e sulle loro conseguenze*, Napoli, 2022, p. 31 ss., 51.

L'art. 3, n. 1, infatti definisce un sistema di intelligenza artificiale 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, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici».

In prima battuta la Commissione europea aveva ipotizzato una definizione flessibile di “sistema di IA” come di «un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono».

Successivamente, però, tale nozione è stata rimeditata al fine di conformarsi alla definizione desumibile dalla Raccomandazione predisposta dall'OCSE, anch'essa, a propria volta, modificata nel corso del 2023²¹. Secondo questa raccomandazione, un sistema di intelligenza artificiale «is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment»²².

Una definizione assolutamente identica è stata poi adottata, nel medesimo lasso temporale, dagli Stati membri del Consiglio d'Europa nella Convenzione quadro sull'intelligenza artificiale e i diritti umani, la democrazia e lo Stato di diritto²³.

Volendo identificare il punto fondamentale nelle definizioni adottate dal regolamento, dalla Raccomandazione dell'OCSE e, infine, dalla Convenzione del Consiglio d'Europa, potremmo affermare che è la capacità di imparare che contraddistingue un sistema di IA. Questo, in definitiva, ci riporta alla mente le parole di Alan Turing che già nel 1950, ipotizzava che in futuro prossimo sarebbero stati inventati dei calcolatori in grado di apprendere e di assumere decisioni conseguenti, quasi come se stessero realmente pensando e non semplicemente mettendo in ordine i dati desunti dall'osservazione della realtà²⁴.

In tal modo il legislatore europeo ha voluto ricomprendere in tale definizione anche la c.d. intelligenza artificiale generativa, ovvero quella che muovendo da dati immessi durante

²¹ Sull'evoluzione della nozione di IA cfr. G. PROIETTI, *Definire l'indefinibile? i sistemi di intelligenza artificiale alla ricerca di un inquadramento sistematico*, in *Contratto e impresa*, 2024, p. 882 ss., 901 ss.

²² OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449. La raccomandazione è stata adottata il 22 maggio 2019. Tale versione è stata modificata in data 8 novembre 2023 e successivamente approvata il 3 maggio 2024.

²³ CM(2024)52-final - Committee on Artificial Intelligence (CAI) - Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, 17/05/2024. Il testo del Rapporto esplicativo, al punto 23, riporta infatti che la definizione adottata dalla Convenzione è ripresa da quella della Raccomandazione dell'OCSE, come modificata nel corso del 2023 («The definition of an artificial intelligence system prescribed in this Article is drawn from the latest revised definition adopted by the OECD on 8 November 2023»). Per il rapporto v. il documento: CM(2024)52-addfinal - Committee on Artificial Intelligence (CAI) - Framework Convention on artificial intelligence, human rights, democracy and the rule of law - Explanatory Report. Tutti i documenti riportati in questa nota sono reperibili sul sito del Consiglio d'Europa (www.coe.int). Per un primo commento cfr. F. SEATZU, *Assessing the Council of Europe's AI Convention: Challenges and Prospects for Protecting Human Rights and Democracy in the Age of AI*, in *Studi sull'integrazione europea*, 2025, p. 9 ss.

²⁴ A.M. TURING, *Macchine calcolatrici e intelligenza*, trad. it. a cura di D. Mauri, Torino, 2025 (il testo originale è del 1950): «Io credo che la domanda “le macchine sono in grado di pensare?”, sia troppo insensata perché valga la pena discuterne. E tuttavia, credo anche che alla fine di questo secolo l'uso delle parole e l'opinione diffusa delle persone colte avranno subito un cambiamento tale che si potrà parlare di macchine che pensano senza aspettarsi di essere contraddetti».

il periodo di addestramento è, a propria volta, capace di reperire le risorse sul web e di elaborare dei testi autonomi²⁵.

Nonostante la definizione di intelligenza artificiale sia estremamente comprensiva, il regolamento prevede talune esclusioni dal proprio ambito di applicazione, in funzione di specifici usi di tale tecnologia. Le principali di dette esclusioni sono le seguenti: *i)* i sistemi di IA sviluppati o utilizzati per scopi esclusivamente *militari o di difesa*; *ii)* i sistemi di IA utilizzati solo a scopo di *ricerca e innovazione*; *iii)* nel caso in cui l'intelligenza artificiale venga utilizzata nel corso di un'*attività personale* a carattere non professionale. Ognuna di queste, evidentemente, risponde ad una ratio specifica. Quella per scopi militari dipende dalla circostanza che il regolamento impone una serie di requisiti di trasparenza e di responsabilità in capo ai fornitori ovvero agli utilizzatori, che sono incompatibili con lo sviluppo delle tecnologie in ambito bellico. D'altro canto, le attività militari si svolgono essenzialmente all'interno di una cornice dettata dal diritto internazionale e, più in particolare, del diritto dei conflitti armati; con la conseguenza che un regolamento europeo mal si presta a circoscriverne l'impiego della tecnologia in esame in tale ambito²⁶. L'eccezione per scopo di ricerca è dovuta al fatto che l'attività scientifica mira esclusivamente a porre in essere sistemi più evoluti: essa stessa ha bisogno di una certa libertà di azione che evidentemente non si concilia con prescrizioni così stringenti quali quelle contenute nel regolamento²⁷. Infine, l'esclusione dei sistemi di IA per meri motivi personali si giustifica per il fatto che questi sistemi hanno un impiego circoscritto all'utilizzatore: non essendo destinati ad uscire dalla sfera di discrezionalità di costui, essi non possono comunque produrre effetti di larga scala, né provocare impatti rilevanti sul godimento dei diritti fondamentali da parte dei terzi.

Infine, per concludere l'introduzione del regolamento, è opportuno chiarire chi ne sono i principali destinatari ovvero i sono i soggetti tenuti ad osservare gli obblighi prescritti dal regolamento stesso²⁸. Il regolamento si applica: *a)* ai fornitori che immettono sul mercato o mettono in servizio sistemi di IA nell'Unione, indipendentemente dal fatto che siano stabiliti nell'unione o in un paese terzo; *b)* agli operatori di sistemi di IA che hanno il loro luogo di stabilimento o che sono ubicati nell'Unione; *c)* ai fornitori e agli operatori di sistemi di IA situati in un paese terzo, laddove l'*output* prodotto dal sistema sia utilizzato nell'Unione; *d)* agli importatori e ai distributori dei sistemi di IA.

Definendo in tal modo il proprio ambito di applicazione, il legislatore dell'Unione europea rende manifesta l'aspirazione a porre in essere delle regole "globali"²⁹ ovvero a

²⁵ Per questi sistemi, la versione finale del regolamento ha previsto la regolamentazione dei modelli di IA per finalità generali, sui quali v., *infra*, nota 58.

²⁶ Considerando n. 24, reg. 2024/1689: «Per quanto riguarda gli scopi militari e di difesa, tale esclusione è giustificata sia dall'articolo 4, paragrafo 2, TUE sia dalle specificità della politica di difesa comune degli Stati membri e dell'Unione di cui al titolo V, capo 2, TUE che sono soggette al diritto internazionale pubblico, che costituisce pertanto il quadro giuridico più appropriato per la regolamentazione dei sistemi di IA nel contesto dell'uso letale della forza e di altri sistemi di IA nel contesto delle attività militari e di difesa».

²⁷ Considerando n. 25, reg. 2024/1689: «Il presente regolamento dovrebbe sostenere l'innovazione, rispettare la libertà della scienza e non dovrebbe pregiudicare le attività di ricerca e sviluppo. È pertanto necessario escludere dal suo ambito di applicazione i sistemi e i modelli di IA specificamente sviluppati e messi in servizio al solo scopo di ricerca e sviluppo scientifici».

²⁸ Sul tema v., più diffusamente, L. GATT, M.T. LO CONTE, M.E. MAZZARELLA, *L'ambito di applicazione, soggettivo e oggettivo, dell'AI Act*, in U. RUFFOLO (a cura di), *AI Act. La regolamentazione europea dell'intelligenza artificiale*, Roma, 2025, p. 49 ss.

²⁹ Obiettivo abitualmente denominato come "sovranità digitale esterna". V. M. CARTA, *Il Regolamento UE sull'Intelligenza Artificiale*, cit., p. 190; G. FINOCCHIARO, *I molteplici obiettivi perseguiti dal regolamento*, in O.

stabilire degli standard che hanno vocazione ad applicarsi anche a soggetti situati in Stati terzi, ma che comunque producono sistemi di intelligenza artificiale destinati ad essere immessi nel mercato interno. Si tratta, evidentemente, di uno scopo non sempre chiaramente esplicitato negli atti che hanno preceduto l'adozione del regolamento, ma che risulta evidentemente ben presente e che, appunto, consiste nell'aspirazione a configurare l'Unione europea come un leader normativo e tecnologico in materia, in grado di definire gli standard globali per l'impiego dell'intelligenza artificiale³⁰.

4. La piramide del rischio

Per contemperare il corretto funzionamento del mercato interno con le esigenze di sicurezza dei sistemi di intelligenza artificiale, il regolamento si fonda su una sorta di «tassonomia del rischio». L'AI Act adotta, infatti, un approccio cosiddetto “*risk-based*”, nel senso che maggiore è il rischio che discende dall'immissione in commercio di un determinato sistema di intelligenza artificiale, più rigorose sono le previsioni che i soggetti coinvolti e, soprattutto i fornitori, devono osservare.

In questa scelta è, peraltro, ravvisabile uno dei limiti di operatività del regolamento. Il sistema finisce, infatti, per assumere una certa rigidità³¹. L'inserimento dell'uno e dell'altro livello rischio è infatti determinato da una scelta a monte effettuata dalla Commissione sulla base delle tipologie di sistemi esistenti e della reportistica degli incidenti in relazione a ciascun utilizzo. Data la rapidità dell'evoluzione tecnologica è inevitabile che sorgano nuovi sistemi capaci di svolgere funzioni ulteriori, che risulteranno difficilmente inquadrabili in una delle categorie regolamentate³².

In base a questo approccio è possibile rintracciare quattro diversi livelli di rischio che vanno dal minore al maggiore. Ciascun livello di rischio prevede specifici permessi e/o restrizioni.

Questa “piramide del rischio” comprende al livello più basso un rischio “minimo” o assente, al quale corrisponde un permesso all'immissione in commercio senza restrizioni. In tale ambito, viene tuttavia raccomandata l'adozione di un Codice di Condotta a fornitori ed utilizzatori.

Il secondo livello di rischio è quello dei testi o filmati elaborati da sistemi di intelligenza artificiale (ad es., i *chatbot*), per il quale è previsto un permesso generale alla loro diffusione a condizione che siano rispettati gli obblighi di informazione e trasparenza. Per detti sistemi, infatti, l'obbligo principale consiste, infatti, nel garantire «che i sistemi di IA destinati a interagire direttamente con le persone fisiche s(ia)no progettati e sviluppati in modo tale che

POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI (a cura di), *La disciplina dell'intelligenza artificiale*, cit., p. 101 ss.

³⁰ Si tratta del noto effetto Bruxelles, come teorizzato da A. BRADFORD, *Effetto Bruxelles. Come l'Unione europea regola il mondo*, Milano, 2021; ID., *Digital Empires: The Global Battle to Regulate Technology*, USA, 2023.

³¹ V. G. FINOCCHIARO, *Diritto dell'intelligenza artificiale*, cit., p. 15.

³² Per questo il regolamento, all'art. 97, prevede una delega alla Commissione per adottare un atto delegato ed emendare l'allegato III, concernente i c.d. sistemi ad alto rischio.

le persone fisiche interessate siano informate del fatto di stare interagendo con un sistema di IA»³³.

Al terzo livello corrisponde il c.d. *rischio alto* e l'autorizzazione all'immissione in commercio è concesso solo a condizione che il fornitore abbia rispettato i requisiti dell'AI Act e abbia adeguatamente svolto una valutazione di conformità *ex-ante*.

Infine il quarto e ultimo livello indica un rischio *inaccettabile*: qui siamo nelle ipotesi nelle quali, secondo la Commissione, ricorrono i maggiori rischi per la tutela delle persone ovvero per i diritti fondamentali. I sistemi di intelligenza artificiale che mirano ad introdurre dei sistemi classificati in questo ambito sono, in linea di principio, vietati. Per ragioni espositive conviene iniziare da questi ultimi.

5. *Il rischio inaccettabile*

I sistemi di intelligenza artificiale a rischio inaccettabile sono disciplinati dal solo art. 5. Questa previsione ha eminentemente la funzione di vietare l'immissione in commercio di sistemi di intelligenza artificiale riconducibili nelle categorie menzionate nella disposizione. Anche se, come vedremo tra breve, i divieti sono estremamente circostanziati e suscettibili di svariate deroghe.

Il primo divieto concerne i sistemi che determinano una manipolazione comportamentale cognitiva, eventualmente attraverso lo sfruttamento di specifiche vulnerabilità (quali età, disabilità o situazione socioeconomica) del soggetto³⁴. In questo caso siamo in presenza di un divieto assoluto, non suscettibile di deroghe. Tuttavia, dalla formulazione delle previsioni normative inerenti, emerge chiaramente che il divieto colpisce non tanto la produzione, quanto piuttosto la commercializzazione ovvero l'uso del sistema diretto a distorcere il comportamento di un determinato soggetto («l'immissione sul mercato, la messa in servizio o l'uso»). Inoltre, per incorrere nel divieto occorre che tali tecniche facciano uso di tecniche subliminali («che utilizza tecniche subliminali») ovvero che sfruttino le debolezze specifiche del soggetto («che sfrutta le vulnerabilità di una persona fisica»); e che si prefiggano, quale scopo specifico, di distorcere il comportamento («aventi lo scopo o l'effetto di distorcere materialmente il comportamento di una persona o di un gruppo di persone»). In quarto luogo occorre che la manipolazione abbia l'effetto di indurre una persona o un gruppo di persone a prendere una decisione che le provochi un danno

³³ Art. 50, reg. 2024/1689. In dottrina v. E. LONGO, F. PAOLUCCI, *The Article 50 of the AI Act and the Transparency obligations: The Model and Its Limitations*, in O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI (a cura di), *La disciplina dell'intelligenza artificiale*, cit., p. 275 ss.

³⁴ Art. 5, reg. 2024/1689: «1. Sono vietate le pratiche di IA seguenti: a) l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che utilizza tecniche subliminali che agiscono senza che una persona ne sia consapevole o tecniche volutamente manipolative o ingannevoli aventi lo scopo o l'effetto di distorcere materialmente il comportamento di una persona o di un gruppo di persone, pregiudicando in modo considerevole la loro capacità di prendere una decisione informata, inducendole pertanto a prendere una decisione che non avrebbero altrimenti preso, in un modo che provochi o possa ragionevolmente provocare a tale persona, a un'altra persona o a un gruppo di persone un danno significativo; b) l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che sfrutta le vulnerabilità di una persona fisica o di uno specifico gruppo di persone, dovute all'età, alla disabilità o a una specifica situazione sociale o economica, con l'obiettivo o l'effetto di distorcere materialmente il comportamento di tale persona o di una persona che appartiene a tale gruppo in un modo che provochi o possa ragionevolmente provocare a tale persona o a un'altra persona un danno significativo».

significativo («inducendole ... a prendere una decisione che non avrebbero altrimenti preso, in un modo che provochi o possa ragionevolmente provocare... un danno significativo»).

Sulla base della formulazione delle disposizioni in esame, è agevole affermare che le condizioni previste sono cumulative: ne discende che devono ricorrere tutte e quattro le condizioni per incorrere nel divieto³⁵. In questo contesto, l'interprete non potrebbe fare ricorso al disposto dell'art. 5, par. 1, lett. a) e b), per vietare l'attività di società quali Neuralink, che elabora supporti per sopperire alle disabilità di specifici soggetti (quali, ad esempio, computer che creano parole dal movimento degli occhi) oppure di chatGPT, che pure può essere utilizzato per creare testi ed immagini molto simili a quelli concepiti dall'intelletto umano³⁶, perché mancano sia dell'elemento oggettivo, sia di quello finalistico: essi non sono, infatti, diretti a indurre un certo comportamento, attraverso tecniche subliminali, per arrecare un danno significativo.

Il secondo divieto concerne il c.d. *social scoring*, ovvero la valutazione del comportamento sociale di determinati soggetti o di talune caratteristiche (in genere, al fine di desumerne elementi per l'accesso a determinati servizi, quali possono essere i servizi socio-assistenziali)³⁷. Tale disposizione rafforza, in definitiva, la previsione già contenuta nell'art. 22 GDPR³⁸. Questo articolo, come è noto, stabilisce il diritto dell'interessato a non essere sottoposto a trattamento automatizzato, a meno che non vi abbia acconsentito ovvero il trattamento sia previsto da una norma nazionale o di diritto dell'Unione. Sia il divieto di *social scoring* previsto dal regolamento sull'intelligenza artificiale, sia il diritto a non essere sottoposto ad un trattamento automatizzato disciplinato dal GDPR richiedono che il trattamento sia produttivo di un danno per l'interessato. La Corte di giustizia, nella nota sentenza *Schufa* ha stabilito che è sufficiente la mera eventualità per l'interessato di ricevere un trattamento

³⁵ Comunicazione della Commissione europea, *Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)*, Brussels, 29.7.2025, C(2025) 5052 final, p. 18, punto 61: «For the prohibition to apply, all four conditions must be simultaneously fulfilled and there must be a plausible causal link between the techniques deployed, the material distortion of the behaviour of the person, and the significant harm that has resulted or is reasonably likely to result from that behaviour» (corsivo aggiunto).

³⁶ V. F. DONATI, *Le pratiche di AI vietate*, in U. RUFFOLO (a cura di), *AI Act*, cit., p. 119 ss., 121; F.P. LEVANTINO, I NERONI REZENDE, *Rischio inaccettabile: usi proibiti*, in O. POLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI (a cura di), *La disciplina dell'intelligenza artificiale*, cit., p. 159 ss., 163.

³⁷ Art. 5, par. 1, lett. c), reg. 2024/1689: «l'immissione sul mercato, la messa in servizio o l'uso di sistemi di IA per la valutazione o la classificazione delle persone fisiche o di gruppi di persone per un determinato periodo di tempo sulla base del loro comportamento sociale o di caratteristiche personali o della personalità note, inferite o previste, in cui il punteggio sociale così ottenuto comporti il verificarsi di uno o di entrambi gli scenari seguenti: i) un trattamento pregiudizievole o sfavorevole di determinate persone fisiche o di gruppi di persone in contesti sociali che non sono collegati ai contesti in cui i dati sono stati originariamente generati o raccolti; ii) un trattamento pregiudizievole o sfavorevole di determinate persone fisiche o di gruppi di persone che sia ingiustificato o sproporzionato rispetto al loro comportamento sociale o alla sua gravità».

³⁸ 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). In dottrina, v., anche per ulteriori citazioni; F. PIZZETTI, *Privacy e il diritto europeo alla protezione dei dati personali*, vol. I (*Dalla dir. 95/46 al nuovo regolamento europeo*) e vol. II (*Il regolamento europeo 2016/679*), Torino, 2016; G. RESTA, V. ZENO ZENCHOVICH, *Volontà e consenso nella fruizione dei servizi in rete*, in *Riv. trim. dir. e proc. civ.*, 2018, p. 411 ss.; V. CUFFARO, *Il diritto europeo sul trattamento dei dati personali*, in *Contratto e impresa*, 2018, 3, p. 1098 ss.; C. CALIFANO, L. COLAPIETRO (a cura di), *Innovazione tecnologica e persona. Il diritto alla protezione dei dati personali nel regolamento UE 2016/679*, Napoli, 2018; N. ZORZI GALGANO (a cura di), *Persona e mercato dei dati. Riflessioni sul GDPR*, Milano, 2019; V. CUFFARO, R. D'ORAZIO, V. RICCIUTO (a cura di), *I dati personali nel diritto europeo*, Torino, 2019; G. FINOCCHIARO, *La protezione dei dati personali in Italia. Regolamento UE 2016/679 e d.lgs. 10 agosto 2018, n. 101*, Bologna, 2019.

deteriore a seguito del trattamento automatizzato³⁹. Se trasponiamo i principi desumibili dall'interpretazione che la Corte di giustizia ha dato dell'art. 22 del regolamento sulla protezione dei dati personali al regolamento sull'intelligenza artificiale, ne discende che, a differenza della prima fattispecie prevista dall'art. 5 (trattamento subliminale per distorcere il comportamento), in questo caso il danno può anche essere meramente eventuale: ovvero discendere dalla circostanza che, per effetto degli elementi utilizzati dal sistema di intelligenza artificiale, ne possa ipoteticamente derivare la privazione di una possibilità per l'interessato (quale può essere, per l'appunto, il mancato accesso ad un contributo sociale o ad un alloggio a costi contenuti, ecc.).

La terza ipotesi concerne la c.d. polizia predittiva ovvero l'uso dei sistemi di intelligenza artificiale «al fine di valutare o prevedere il rischio che una persona fisica commetta un reato, unicamente sulla base della profilazione di una persona fisica o della valutazione dei tratti e delle caratteristiche della personalità»⁴⁰. Si tratta, in altri termini, del ricorso ai sistemi di intelligenza artificiale al fine di impedire o prevenire la commissione di reati ovvero calcolare una pena detentiva. Il divieto trae origine dal noto caso COMPAS (*Correctional offender management profiling for alternative sanctions*), nel quale una corte americana⁴¹ determinò la pena detentiva facendo ricorso ad un sistema di intelligenza artificiale artatamente capace di valutare la propensione ad incorrere in recidiva sulla base di determinati parametri, quali la storia pregressa, l'ambiente sociale di provenienza, la condizione economica, il livello di istruzione. In altri termini: è evidente che una valutazione siffatta presenti un elevato rischio di discriminazioni fondate su preconcetti statistici.

Tale disposizione va coordinata con quella contenuta al punto 8 dell'Allegato III, del regolamento il quale ricomprende, tra i sistemi ad alto rischio, quelli che fanno uso dei sistemi di intelligenza artificiale nell'ambito di un processo «per assistere un'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti, o a essere utilizzati in modo analogo nella risoluzione alternativa delle controversie». Sulla base del combinato disposto dell'art. 5, par. 1, lett. c, e dell'Allegato III, sembrerebbe possibile adombrare la seguente conclusione: mentre è possibile il ricorso ad un sistema automatizzato per ricercare fatti o elementi ermeneutici, lo stesso non sarebbe ammissibile per calcolare la pena o, addirittura, scrivere una sentenza⁴². In questi casi, infatti,

³⁹ Corte di giustizia, 7 dicembre 2023, causa C- 634/21, *OQ c. Land Hessen e SCHUFA Holding AG*, ECLI:EU:C:2023:957, spec. punti 48-50.

⁴⁰ Art. 5, par. 1, lett. d), reg. 2024/1689.

⁴¹ Corte Suprema del Wisconsin, 13 July 2016, *State v. Loomis*, Case no. 2015AP157-CR (la decisione è consultabile al sito <https://law.justia.com/>). Il caso *Loomis* è talmente conosciuto da venire richiamato anche in una sentenza del Consiglio di Stato (Cons. Stato, sez. VI, 13 dicembre 2019, n. 8474, par. 15.2, seppure in relazione all'uso dei procedimenti automatizzati per l'immissione in ruolo dei docenti). In dottrina v. K. FREEMAN, *Algorithmic injustice: how the Wisconsin Supreme Court failed to protect due process rights in State v. Loomis*, in *North Carolina Journal of Law and Technology*, 2018, p. 75 ss.; F. LAGIOIA, G. SARTOR, *Il sistema COMPAS: algoritmi, previsioni, iniquità*, in U. RUFFOLO (a cura di), *XXVI Lezioni di diritto dell'intelligenza artificiale*, Torino, Giappichelli, 2021, p. 226 e ss.; G. GALLONE, *Riserva di umanità e funzioni amministrative. Indagine sui limiti dell'automazione decisionale tra procedimento e processo*, Padova, 2023, 117 s.; C. PAPPALARDO, *Le applicazioni e le pratiche proibite di IA*, in G. CASSANO, E.M. TRIPODI (a cura di), *Il regolamento europeo sull'intelligenza artificiale. Commento al Reg. UE n. 1689/2024*, Sant'Arcangelo di Romagna, 2024, p. 441 ss., 460.

⁴² In tal senso v. anche la raccomandazione approvata dal plenum del Consiglio superiore della magistratura, in data 8 ottobre 2025, 118/VV/2024: «Riaffermare la centralità dell'esperienza processuale. Il processo e la valutazione giudiziale sono un'esperienza umana insostituibile, fondata su presenza, oralità, immediatezza percettiva e contraddittorio in uno spazio condiviso. Il Consiglio si adopererà, pertanto, per impedire la “despazializzazione del giudizio”, ossia la riduzione del processo a mera interazione digitale, riaffermando che

dovrebbe essere sempre necessario l'intervento umano, per non incorrere nel divieto assoluto di immissione in commercio e di impiego⁴³.

Il quarto divieto riguarda i sistemi di IA che creano immagini ovvero ampliano banche dati mediante tecniche di «*scraping* non mirato di immagini facciali da internet o da filmati di telecamere a circuito chiuso» (art. 5, par. 1, lett. e). Le tecniche di *scraping* consistono nel raccogliere immagini da telecamere ovvero da siti internet e *social network*, al fine di modificarle ovvero di ampliare i *data base*. La ratio del divieto risiede nel fatto che queste tecniche possono facilmente violare la *privacy* degli individui e «accresce il senso di sorveglianza»⁴⁴ da parte della collettività.

Tale divieto va tuttavia distinto da quello concernente la «categorizzazione biometrica» in tempo reale, per dedurre dati sensibili e la «identificazione biometrica», che costituiscono le ipotesi previste dalle lettere h) e g) della medesima disposizione⁴⁵. In questo caso si possono, quindi, determinare due situazioni parallele dal punto di vista delle tecniche impiegate, ma assai diverse quanto ad estensione del divieto. La differenza risiede nel fatto che, mentre nel caso di *scraping* le immagini di un individuo vengono utilizzate per inferirne determinati caratteri o l'appartenenza a specifici gruppi (quali, in via esemplificativa, partiti politici, credenze religiose, orientamenti sessuali), nel secondo caso le tecniche di categorizzazione biometrica sono utilizzate al mero fine di riconoscere una persona (che magari sta volutamente celando la propria identità)⁴⁶.

È poi necessario introdurre un'ulteriore distinzione tra le due ipotesi prese in considerazione dalle lettere h) e g). Mentre nel primo caso (la categorizzazione biometrica) il divieto è assoluto, perché si scontra soprattutto con il divieto di trattamento di dati personali

la decisione nasce da una relazione viva tra giudice, parti e prove, in un tempo e in uno spazio che ne garantiscano serietà, pubblicità e responsabilità».

⁴³ Sul tema v. R.E. KOSTORIS, *Intelligenza artificiale, strumenti predittivi e processo penale*, in *Cass. pen.*, 2024, p. 1642 ss.; U. RUFFOLO, *Sentenza robotica e giustizia predittiva*, in U. RUFFOLO, A. AMIDEI, *Diritto dell'intelligenza artificiale*, cit., p. 15 ss.; E. PIETROCARLO, *Predictive policing e giustizia predittiva nel Regolamento europeo sull'IA*, in U. RUFFOLO (a cura di), *AI Act*, cit., p. 231 ss.

⁴⁴ Considerando n. 43, reg. 2024/1689.

⁴⁵ V. art. 5, par. 1, rispettivamente lett. g) e h): «g) l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di sistemi di *categorizzazione biometrica* che classificano individualmente le persone fisiche sulla base dei loro dati biometrici per trarre deduzioni o inferenze in merito a razza, opinioni politiche, appartenenza sindacale, convinzioni religiose o filosofiche, vita sessuale o orientamento sessuale; tale divieto non riguarda l'etichettatura o il filtraggio di set di dati biometrici acquisiti legalmente, come le immagini, sulla base di dati biometrici o della categorizzazione di dati biometrici nel settore delle attività di contrasto; h) l'uso di sistemi di *identificazione biometrica* remota «in tempo reale» in spazi accessibili al pubblico a fini di attività di contrasto a meno che, e nella misura in cui, tale uso sia strettamente necessario per uno degli obiettivi seguenti: i) la ricerca mirata di specifiche vittime di sottrazione, tratta di esseri umani o sfruttamento sessuale di esseri umani, nonché la ricerca di persone scomparse; ii) la prevenzione di una minaccia specifica, sostanziale e imminente per la vita o l'incolumità fisica delle persone fisiche o di una minaccia reale e attuale o reale e prevedibile di un attacco terroristico; iii) la localizzazione o l'identificazione di una persona sospettata di aver commesso un reato, ai fini dello svolgimento di un'indagine penale, o dell'esercizio di un'azione penale o dell'esecuzione di una sanzione penale per i reati di cui all'allegato II, punibile nello Stato membro interessato con una pena o una misura di sicurezza privativa della libertà della durata massima di almeno quattro anni» (corsivi aggiunti).

⁴⁶ Per la distinzione tra categorizzazione e individuazione biometrica attraverso tecniche di riconoscimento facciale v. Art. 29, Data Protection Working Party, Opinion 02/2012 on facial recognition in online and mobile services, 00727/12/EN, WP 192, Brussels, 22 March 2012; S. DEL GATTO, *La governance delle nuove tecnologie tra tentativi di regolazione e istanze di self regulation. il caso del riconoscimento facciale*, in *Riv. ital. dir. pubbl. com.*, 2023, p. 37 ss., 41 s.; N. MENÉNDEZ GONZÁLEZ, G. MOBILIO, *Between Prohibited Risks and High Risks: The Regulation of Facial Recognition Technology*, in O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI (a cura di), *La disciplina dell'intelligenza artificiale*, cit., p. 199 ss.

che «rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose» di cui all'art. 9 GDPR, nel secondo caso («l'identificazione biometrica») il divieto è, al contrario, derogabile se ricorrono determinate condizioni⁴⁷. In questa seconda ipotesi, infatti, il sistema di intelligenza artificiale interviene solo per agevolare l'individuazione di determinate persone: in questo caso, vista l'esigenza di sicurezza della collettività, il divieto è derogabile se serve per identificare persone che siano collegate (in quanto vittime o sospetti autori) con determinati reati, particolarmente gravi e potenzialmente lesivi dell'ordine pubblico.

Da ultimo, la disposizione in esame vieta il riconoscimento delle emozioni sul luogo di lavoro e negli istituti di istruzione (art. 5, par. 1, lett. f). Sono infatti noti gli impieghi di sistemi di intelligenza artificiale al fine di individuare le emozioni in questi luoghi. Il considerando n. 44, del regolamento 2024/1689 giustifica tale divieto sulla base delle differenze rilevabili tra le varie popolazioni. Si può infatti leggere in detto considerando che: «Sussistono serie preoccupazioni in merito alla base scientifica dei sistemi di IA diretti ad identificare o inferire emozioni, in particolare perchè l'espressione delle emozioni varia notevolmente in base alle culture e alle situazioni e persino in relazione a una stessa persona... Pertanto, i sistemi di IA che identificano o inferiscono emozioni o intenzioni di persone fisiche sulla base dei loro dati biometrici possono portare a risultati discriminatori e possono essere invasivi dei diritti e delle libertà delle persone interessate».

Quindi, la preoccupazione verso i sistemi di riconoscimento delle emozioni, concerne la circostanza che detti sistemi risultano gravemente pregiudizievoli per i diritti fondamentali⁴⁸. Siffatti impieghi dei sistemi di intelligenza artificiale, peraltro diffusi in taluni paesi asiatici (ad esempio in Cina), in quanto si pongono in aperta violazione della privacy e della vita privata degli individui soggetti al controllo e, come tali, si pongono in aperto contrasto con gli articoli 7 e 8 della Carta dei diritti fondamentali.

Tale divieto, peraltro, consente un'eccezione, nella misura in cui il riconoscimento delle emozioni possa produrre effetti positivi per l'individuo, quali la diagnosi e l'individuazione del trattamento sanitario (specie nei casi di malattie mentali) ovvero di sicurezza⁴⁹.

6. I sistemi ad alto rischio

Il terzo livello della piramide è occupato dai sistemi di intelligenza artificiale ad alto rischio. Si tratta, peraltro, della fattispecie alla quale sono destinate la maggior parte delle disposizioni del regolamento europeo.

⁴⁷ L'attività di *web scraping* è stata una delle violazioni commesse dalla società statunitense Clearview AI, la quale scandaglia il web al fine di individuare di individuare elementi biometrici di identificazione, che successivamente trasforma in algoritmi, al fine di procedere all'identificazione dei soggetti, senza acquisire il previo consenso di questi. V. il caso deciso dal Garante per la protezione dei dati personali, Ordinanza ingiunzione nei confronti di Clearview AI - 10 febbraio 2022, 9751362; in dottrina v. N. ZORZI GALGANO, *Il regolamento UE 2024/1689 del 13 giugno 2024 sul c.d. alto rischio inaccettabile: le pratiche inerenti sistemi di intelligenza artificiale vietate dal legislatore europeo*, in *Contratto e impresa*, 2025, p. 41 ss., 71 s.

⁴⁸ E. M. RENIERIS, *Beyond Data. Reclaiming Human Rights at the Dawn of the Metaverse*, MIT Press, 2023, 102, riporta che in Cina se ne fa uso anche per accertare se gli studenti in età scolare siano effettivamente felici.

⁴⁹ Non rientrano evidentemente nel divieto i sistemi di controllo che non servono per desumere le emozioni del soggetto. Le linee guida elaborate dalla Commissione, al riguardo, riporta quale esempio il caso di impiego di sistemi di IA che seguono gli occhi degli studenti durante un esame, per evitare casi di falso ovvero di plagio (*Commission Guidelines on prohibited artificial intelligence practices*, cit., p. 86 s.).

Ai sensi dell'art. 6 ricorre un'ipotesi di sistema di IA ad alto rischio in presenza di due situazioni eterogenee.

Innanzitutto, allorché il sistema di intelligenza artificiale costituisce uno dei prodotti (ovvero una componente di questi) oggetto della normativa armonizzata a livello dell'Unione e ricompresa nell'allegato I al regolamento. Detto allegato è estremamente ampio e include un lungo elenco di normative di armonizzazione. È impossibile, nell'economia di questo scritto, elencarle se non in maniera sintetica. Le principali riguardano: la sicurezza dei giocattoli; gli ascensori; le apparecchiature radio; gli impianti a fune; i dispositivi di protezione individuale; i dispositivi medici; la sicurezza dell'aviazione civile; i mezzi a due o tre ruote; l'equipaggiamento marittimo; l'interoperabilità del sistema ferroviario⁵⁰.

Nel caso in cui il sistema di IA, pur non rientrando nell'ambito della normativa armonizzata elencata nell'allegato I, è destinato a particolari fini, allora l'interprete deve verificare se l'ipotesi rientri comunque in una di quelle previste dall'allegato III. Detto allegato menziona diversi impieghi dei sistemi di intelligenza artificiale, che riguardano, rispettivamente: l'identificazione biometrica a posteriori; il riconoscimento delle emozioni; la formazione professionale; la valutazione e l'apprendimento; la selezione personale; l'accesso ai servizi essenziali; il controllo delle frontiere; l'uso, da parte dell'autorità giudiziaria per la ricerca dei fatti o del diritto ovvero quale strumento nella risoluzione alternativa delle controversie⁵¹.

In definitiva, la Commissione europea ha optato per una soluzione definitiva rigida dei sistemi ad alto rischio. Essa si è, infatti, basata su meri dati statistici ovvero sul calcolo delle probabilità che si verifichino degli incidenti nell'uso di detti sistemi; e, sulla base di detto riscontro, ne ha poi previsto l'inclusione nella categoria in esame⁵². Rispetto a detta rigidità sono, tuttavia, ammessi due temperamenti. Il primo è che la Commissione può modificare l'Allegato III, attraverso atti delegati, che può esercitare «aggiungendo o modificando i casi d'uso dei sistemi di IA ad alto rischio»⁵³. In secondo luogo, un sistema di IA, pur rientrando in uno degli usi di cui all'Allegato III, non è considerato ad alto rischio se lo stesso è utilizzato solo per svolgere «un compito procedurale limitato» ovvero a «migliorare un'attività umana precedentemente completata», senza sostituire o influenzare la valutazione umana⁵⁴.

Gli obblighi principali che incombono sui fornitori dei sistemi ad alto rischio sono disciplinati dagli articoli 9 a 15. Questi, sulla falsariga delle norme sulla sicurezza dei prodotti, comprendono sia obblighi di documentazione, tracciabilità e trasparenza; sia obblighi che attengono ai dati che devono essere utilizzati per addestrare il sistema; sia, infine, obblighi di

⁵⁰ Omettiamo di riportare in nota l'elenco (piuttosto esteso) dei singoli atti normativi compresi nell'Allegato I. Ciò che conta è siano disposizioni di armonizzazione recenti: quelle correntemente definite come appartenenti alla nuova cornice legislativa (NFL – New Legislative Framework). V. F. GALLI, G. CONTISSA, G. SARTOR, *I sistemi ad alto rischio*, in U. RUFFOLO (a cura di), *AI Act. La regolamentazione europea dell'intelligenza artificiale*, cit., p. 155 ss., 157; T. RODRÍGUEZ DE LAS HERAS BALLEL, *High-Risk AI Systems in the European AI Act*, in O. POLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI (a cura di), *La disciplina dell'intelligenza artificiale*, cit., p. 228 ss., 239.

⁵¹ L'Allegato III, pertanto, non sta legittimando un uso dell'intelligenza artificiale quale sostitutivo delle parti, degli Avvocati ovvero dei Giudici, ma solo quale ausilio allo stesso. V. U. RUFFOLO, *I due volti della giustizia predittiva: decisoria e consultiva*, in U. RUFFOLO (a cura di), *AI Act*, cit., p. 193 ss.; M. BOLOGNESE, *La funzione giurisdizionale nel diritto europeo dell'intelligenza artificiale*, in *rivista.eurojus.it*, 3/2025, p. 207 ss.

⁵² Le ipotesi previste in maniera generale ed astratta dal regolamento, discendono dalla combinazione della probabilità statistica che si verifichi un danno, con la gravità del danno stesso. V. T. RODRÍGUEZ DE LAS HERAS BALLEL, *High-Risk AI Systems in the European AI Act*, cit., p. 237.

⁵³ Art. 7, par. 1, reg. 2024/1689.

⁵⁴ Art. 6, par. 3, reg. 2024/1689.

introdurre sistemi resilienti. In definitiva si tratta di una serie di regole di prudenza, che assolvono eminentemente alla funzione di «ricondere i rischi posti dai sistemi di IA ad alto rischio entro livelli ritenuti “accettabili”»⁵⁵.

Più specificamente gli oneri di documentazione concernono: l'obbligo dei fornitori di mettere in piedi e aggiornare costantemente un sistema di monitoraggio e gestione dei rischi (art. 9); l'obbligo di redigere una documentazione tecnica prima dell'immissione nel mercato (art. 11); l'obbligo di tracciare gli eventi che riguardano il sistema e di conservarne la relativa documentazione durante tutta la vita dello stesso (art. 12); l'obbligo, infine, di garantire che le informazioni siano chiare, trasparenti e facilmente accessibili per gli utilizzatori (art. 13). Detti obblighi sono completati da una disposizione, introdotta solo in occasione dell'ultima versione del regolamento, che concerne il diritto di ogni interessato, in ipotesi di decisione assunta da un utilizzatore attraverso un sistema di intelligenza artificiale ad alto rischio, di agire per ottenere una spiegazione dall'utilizzatore del sistema, circa il «ruolo del sistema di IA nella procedura decisionale e sui principali elementi della decisione adottata» (art. 86).

Per quanto riguarda l'addestramento del sistema, il fornitore deve garantire che questo avvenga solo con dati di alta qualità: questi sono i dati adeguati alla finalità specifica, rispetto ai quali è adeguatamente documentata la prestazione del consenso del titolare, in ossequio alle previsioni del citato regolamento sulla tutela dei dati personali (art. 10).

Il fornitore deve poi garantire che i sistemi ad alto rischio siano «progettati e sviluppati in modo tale da conseguire un adeguato livello di accuratezza, robustezza e cibersecurity e da operare in modo coerente con tali aspetti durante tutto il loro ciclo di vita» (art. 15).

Ulteriori obblighi specifici per i fornitori riguardano quello di apporre il proprio segno distintivo; creare dei punti di contatto; mettere in piedi un sistema di gestione della qualità (art. 17); conservare la documentazione per dieci anni (art. 18); apporre la marcatura CE per indicare la conformità alla normativa unionale (48).

Un obbligo specifico per gli utilizzatori, riguarda, invece, quello di valutare l'impatto del sistema di IA ad alto rischio sui diritti fondamentali (art. 27).

Superiore e di portata tale da contraddistinguere l'approccio di tutto il regolamento ai sistemi ad alto rischio è poi l'obbligo che incombe sia ai fornitori (art. 14), sia agli utilizzatori (art. 26), di garantire la possibilità di una sorveglianza umana in ogni fase del processo di funzionamento. Tale obbligo, che fa il palio con la previsione del GDPR che prevede il diritto di ogni interessato a non essere sottoposto ad una decisione automatizzata (il citato art. 22, del reg. 2016/679), è, in qualche modo, il centro normativo intorno al quale ruotano tutte le previsioni del regolamento e che è atto a caratterizzare l'AI Act come un sistema effettivamente antropocentrico, in quanto diretto a porre al centro la persona umana e a considerare quest'ultima come ultimo ed effettivo responsabile di ogni violazione.

7. La governance dell'IA

Al fine di garantirne una corretta applicazione il legislatore europeo ha previsto l'istituzione di vari organi che svolgono compiti consultivi o di controllo che danno vita a

⁵⁵ R. PETRUSO, G. SMORTO, *Il Regolamento europeo sull'intelligenza artificiale: una prima lettura*, in *Nuova giur. civ. comm.*, 2024, II, p. 989 ss., 996.

una nuova architettura di governance. Tale sistema è estremamente complesso ed articolato⁵⁶. Il quadro generale che ne discende non risulta affatto di agevole lettura, dal momento che prevede frequenti interrelazioni e sovrapposizioni di competenze.

Per comodità espositiva, conviene distinguere il livello europeo da quello nazionale. Per quanto concerne l'ordinamento sovranazionale, il regolamento prevede l'istituzione di ben cinque autorità. Alcune di queste sono già esistenti, ma – per effetto del regolamento – divengono tributarie di nuove funzioni.

La prima delle autorità europee ha compiti esecutivi: si tratta, nella specie, dell'Ufficio per l'intelligenza artificiale, istituito all'interno della Commissione per far rispettare le regole stabilite dal regolamento all'interno dell'Unione europea⁵⁷. Questo ufficio monitora l'efficace attuazione del regolamento e, a tal fine, può: chiedere documentazione sui modelli di IA, valutare la conformità del fornitore dei modelli di IA agli obblighi previsti dal regolamento; richiedere l'accesso al modello stesso⁵⁸.

Vi sono, poi, tre organi che assolvono eminentemente a funzioni consultive. Il primo di questi è il Consiglio per l'intelligenza artificiale (art. 65-66). Esso è composto da rappresentanti degli Stati membri ed è incaricato di fornire consulenza e assistenza alla Commissione e agli Stati membri ai fini di un'applicazione coerente ed efficace del regolamento sull'intelligenza artificiale. Il secondo è un Forum consultivo (art. 67) per i portatori di interessi volto a fornire competenze tecniche al Consiglio per l'intelligenza artificiale e alla stessa Commissione europea. Da ultimo, il regolamento prevede l'istituzione del Gruppo di esperti scientifici indipendenti (art. 68), selezionati dalla Commissione europea per le loro competenze specifiche, a sostegno delle attività di esecuzione, che assistono la

⁵⁶ Sul sistema di *governance* v. M. RABITTI, F. BASSAN, *L'applicazione dell'AI Act in Italia e la tutela del consumatore. Il ruolo delle autorità indipendenti*, in *Federalismi*, n. 30/2024, p. 250 ss.; S. VILLANI, *Il sistema di vigilanza sull'applicazione dell'AI Act: ognun per sé?*, in *Quaderni AISDUE*, Fascicolo speciale n. 2/2024, *La nuova disciplina UE sull'intelligenza artificiale*, p. 1 ss.; C. CALENDIA, *La governance a livello dell'Unione europea. Il Comitato europeo per l'intelligenza artificiale e gli altri organismi di supporto*, in G. CASSANO, E.M. TRIPODI (a cura di), *Il regolamento europeo sull'intelligenza artificiale*, cit., p. 743 ss.; A. PAJNO, *La governance dell'Intelligenza artificiale fra Regolamento europeo e disciplina nazionale*, in U. RUFFOLO (a cura di), *AI Act. La regolamentazione europea dell'intelligenza artificiale*, Roma, 2025, p. 353 ss.; G. NATALE, *il ruolo delle autorità amministrative indipendenti in materia di AI*, ivi, p. 385 ss.

⁵⁷ Art. 64 del regolamento 2024/1689. L'ufficio è stato istituito con decisione della Commissione, prima che venisse licenziato il testo definitivo del regolamento con la decisione C/2024/1459, del 24 gennaio 2024, che istituisce l'ufficio europeo per l'Intelligenza Artificiale, in GUUE, C/2024/1459, del 14 febbraio 2024.

⁵⁸ L'Ufficio per l'IA si occupa anche della supervisione sui modelli di IA per finalità generali (v. considerando 151 e art. 64, del regolamento 2024/1689). Tali sono quei sistemi che siano allenati «su un largo insieme di dati (generalmente usando auto-supervisione ad ampia scala) e che può essere adattato (per esempio, accordato a un ampio insieme di compiti a valle)». In concreto, sarebbero i sistemi di IA c.d. generativa, che hanno una grande flessibilità e una capacità di fornire risposte, elaborare testi e immagini in maniera autonoma, sulla base dei dati elaborati durante il processo di addestramento. Il regolamento prevede obblighi specifici per i fornitori di modelli di IA per finalità generali e obblighi più stringenti per i sistemi di IA che prevedono rischi sistemici: v., rispettivamente, gli articoli 53 e 55. In dottrina v. G. SARTOR, G. DE GREGORIO, G. MUTO, *Large Language Models, Intelligenza artificiale generativa e Artificial Intelligence Act*, in O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI (a cura di), *La disciplina dell'intelligenza artificiale*, cit., p. 248 ss., spec. 254 s.; U. RUFFOLO, *AI Act e AI generativa*, in U. RUFFOLO (a cura di), *AI Act. La regolamentazione europea dell'intelligenza artificiale*, cit., p. 303 ss.; M. UNFER, *L'IA generativa per finalità generali*, in G. CASSANO, E.M. TRIPODI (a cura di), *Il regolamento europeo sull'intelligenza artificiale*, cit., p. 653 ss.; ID., *Segue: ChatGPT e sistemi simili*, ivi, p. 679 ss.; L. CHIMENTI, *L'IA che genera immagini e video*, ivi, p. 687 ss.

Commissione europea e le autorità di vigilanza nazionali soprattutto per quanto concerne i modelli di intelligenza artificiale per finalità generali⁵⁹.

Infine, l'ultimo organo europeo che ha assunto nuove funzioni per effetto del regolamento è il Garante europeo della protezione dei dati, che assolve al ruolo di autorità di vigilanza nei confronti delle istituzioni europee, ipotesi che ricorrerà soprattutto nel caso in cui queste svolgano il ruolo di utilizzatrici di sistemi di intelligenza artificiale (art. 70, par. 9).

Focalizzando ora l'attenzione sul piano dell'ordinamento interno, gli Stati membri devono designare due autorità nazionali competenti: una per la vigilanza del mercato e una per la ricezione delle notifiche (art. 70); l'altra, per garantire il rispetto dei diritti fondamentali (art. 77).

Sul piano dell'ordinamento italiano⁶⁰, le autorità competenti per garantire il rispetto dei diritti fondamentali sono l'AGCOM per quanto concerne l'informazione, i contenuti (testuali, fotografici, audiovisivi e radiofonici), nonché il diritto d'autore; il garante protezione dati personali per la protezione dei dati personali; i giudici nazionali.

Inoltre la legge 132/2025⁶¹, in attuazione del regolamento, ha così ripartito i compiti alle varie strutture interne. Innanzitutto, il legislatore ha istituito una struttura presso la Presidenza del Consiglio dei ministri, che deve provvedere alla predisposizione della strategia nazionale per l'intelligenza artificiale. Questa serve per favorire «la collaborazione tra le amministrazioni pubbliche e i soggetti privati relativamente allo sviluppo e all'adozione di sistemi di intelligenza artificiale, coordina(re) l'attività della pubblica amministrazione in materia, promuovere la ricerca e la diffusione della conoscenza in materia di intelligenza artificiale e indirizza le misure e gli incentivi finalizzati allo sviluppo imprenditoriale e industriale dell'intelligenza artificiale» (art. 19, secondo comma, l. 132/2025).

La struttura presso la Presidenza del Consiglio dei ministri «provvede al coordinamento e al monitoraggio dell'attuazione della strategia (...), avvalendosi dell'Agenzia per l'Italia digitale, d'intesa, per gli aspetti di competenza, con l'Agenzia per la cybersicurezza nazionale e sentiti la Banca d'Italia, la Commissione nazionale per le società e la borsa (CONSOB) e l'Istituto per la vigilanza sulle assicurazioni (IVASS) in qualità di autorità di vigilanza del mercato» (art. 19, quarto comma, l. 132/2025).

In sintesi: la principale autorità di vigilanza nazionale è l'ACN (Agenzia Nazionale per la Cybersicurezza), la quale si occupa della vigilanza, nonché delle attività ispettive e

⁵⁹ In questo ambito il Gruppo di esperti ha anche limitati poteri esecutivi. Il considerando 163 prevede, infatti che: «Al fine di integrare i sistemi di governance per i modelli di IA per finalità generali, il gruppo di esperti scientifici dovrebbe sostenere le attività di monitoraggio dell'ufficio per l'IA e può, in alcuni casi, fornire segnalazioni qualificate all'ufficio per l'IA che avviano attività di follow-up quali indagini. Ciò dovrebbe valere quando il gruppo di esperti scientifici ha motivo di sospettare che un modello di IA per finalità generali rappresenti un rischio concreto e identificabile a livello dell'Unione. Inoltre, ciò dovrebbe verificarsi nel caso in cui il gruppo di esperti scientifici abbia motivo di sospettare che un modello di IA per finalità generali soddisfa i criteri che comporterebbero una classificazione come modello di IA per finalità generali con rischio sistemico. Per dotare il gruppo di esperti scientifici delle informazioni necessarie per lo svolgimento di tali compiti, è opportuno prevedere un meccanismo in base al quale il gruppo di esperti scientifici può chiedere alla Commissione di esigere da un fornitore documentazione o informazioni».

⁶⁰ Sul sistema italiano v. E. GUARNACCIA, *Le autorità nazionali competenti*, in G. CASSANO, E.M. TRIPODI (a cura di), *Il regolamento europeo sull'intelligenza artificiale*, cit., p. 743 ss.; R. LATTANZI, *Al Act e autorità di protezione dei dati personali*, in O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI (a cura di), *La disciplina dell'intelligenza artificiale*, cit., p. 367 ss.; A. DE TOMMASO, M.L. MARINELLO, *Il ruolo di AGCOM nella regolazione dell'intelligenza artificiale*, ivi, p. 399 ss.

⁶¹ L. 23 settembre 2025, n. 132, Disposizioni e deleghe al Governo in materia di intelligenza artificiale, in GU n. 223, del 25 settembre 2025.

sanzionatorie dei sistemi di intelligenza artificiale, oltre, ovviamente, a curare i profili della sicurezza. Inoltre la legge ha istituito l'AgID, l'Agenzia per l'Italia Digitale, quale autorità di notifica e per effettuare i compiti di valutazione, accreditamento e monitoraggio dei soggetti incaricati di verificare la conformità dei sistemi di intelligenza artificiale (art. 20, l. 132/2025).

Vi è infine anche un Comitato per il coordinamento presso Presidenza Consiglio dei Ministri, presieduto dal Presidente del Consiglio dei ministri, che «svolge funzioni di coordinamento dell'azione di indirizzo e di promozione delle attività di ricerca, di sperimentazione, di sviluppo, di adozione e di applicazione di sistemi e modelli di intelligenza artificiale svolte da enti e organismi nazionali pubblici o privati soggetti a vigilanza o destinatari di finanziamento pubblico, ivi comprese le fondazioni pubbliche o private vigilate o finanziate dallo Stato, che operano nel campo dell'innovazione digitale e dell'intelligenza artificiale».

In conclusione, vi è una congerie di autorità sia europee, sia nazionali, ognuna teoricamente competente per compiti specifici, ma che in realtà dovrà necessariamente coordinarsi e tenere conto delle attività poste in essere dalle altre autorità.

A questo punto viene spontaneo chiedersi sulla base di quali principi debba operarsi il coordinamento. Non può infatti liquidarsi, in maniera semplicistica, il tema del coordinamento, limitandoci a stabilire la prevalenza delle autorità europee su quelle nazionali, come si verifica, ad esempio in materia di concorrenza, nei rapporti tra la Commissione europea e le autorità nazionali incaricate del controllo sull'applicazione dei divieti in materia antitrust. Il principio della prevalenza del diritto dell'Unione europea, infatti, non risolve il problema del coordinamento, perché i compiti delle varie autorità coinvolte nell'applicazione del regolamento sull'intelligenza artificiale si assestano su piani paralleli, tra di loro non coincidenti: mentre quelle europee esercitano funzioni esecutive e consultive; quelle nazionali esercitano compiti di vigilanza e di controllo. Non ha pertanto senso a ribadire la prevalenza dell'ordinamento sovranazionale su quello nazionale, perché questo principio non apporta elementi utili, stante le evidenti diversità di ruoli.

L'unico caso nel quale la Corte di giustizia ha affrontato il problema del coordinamento tra autorità tributarie di funzioni differenti si è posto nella sentenza *Meta c. Bundeskartellamt*. In detta occasione la Corte di Lussemburgo affermò l'esistenza di un obbligo di coordinamento tra le autorità. Si può, infatti, leggere in detta sentenza che: «se è vero che... né il GDPR né alcun altro strumento del diritto dell'Unione prevedono norme specifiche a tal riguardo, ciò non toglie che... quando applicano il GDPR, le diverse autorità nazionali coinvolte sono tutte vincolate dal principio di leale cooperazione sancito all'articolo 4, paragrafo 3, TUE. ...in forza di tale principio, nelle materie rientranti nel diritto dell'Unione, gli Stati membri, *ivi incluse le loro autorità amministrative, devono rispettarci ed assistersi reciprocamente nell'adempimento dei compiti derivanti dai Trattati*, adottare ogni misura atta ad assicurare l'esecuzione degli obblighi conseguenti, in particolare, agli atti delle istituzioni dell'Unione, nonché astenersi da qualsiasi misura che rischi di mettere in pericolo la realizzazione degli obiettivi dell'Unione»⁶².

⁶² Corte di giustizia, 23 luglio 2023, causa C-252/21, *Meta c. Bundeskartellamt*, ECLI:EU:C:2023:537, punto 53. Sul caso v., anche per ulteriori riferimenti, F. BATTAGLIA, *Il consumatore digitale nel diritto dell'Unione europea*, Napoli, 2023, p. 267 ss.; ID., *La sentenza Meta platforms. Riflessioni in materia di valore dei dati e libera espressione del consenso*, in *OIDU*, 2023, p. 723 ss.; P. MANZINI, *Antitrust e privacy: la strana coppia*, in *Quaderni AISDUE*, 2023, p. 196 ss.; P.J. VAN DE WAERDT, *Meta v Bundeskartellamt: Something Old, Something New*, in *European Papers*, 2023, p. 1077 ss.; A. LICASTRO, *Meta Platforms Inc., già Facebook Inc. v. Bundeskartellamt: la Corte di Giustizia dell'Unione Europea apre (finalmente) all'integrazione fra diritto antitrust e data protection*, in *Rivista della Regolazione dei mercati*, 2023, p. 503 ss.;

In sintesi, la Corte estese il principio di leale cooperazione, che per espressa previsione è destinato ad operare tra Stati membri ed istituzioni europee ovvero nei rapporti orizzontali tra queste ultime, anche ai rapporti intercorrenti tra autorità amministrative nazionali, allorché esse applicano il diritto dell'Unione europea.

Occorre tuttavia chiedersi, se un siffatto principio sia sufficiente nella materia oggetto delle riflessioni che stiamo svolgendo in questo paragrafo. È infatti ovvio che un conto è coordinare l'attività di due autorità che si occupano di materie eterogenee, come si verificava, per l'appunto nel caso *Meta* (rispettivamente: l'autorità garante della concorrenza e quella dedita alla protezione dei dati personali); mentre cosa ben diversa è coordinare l'attività di una pluralità di enti diversi con competenze non definite in maniera precisa e che appaiono parzialmente sovrapponibili. Ci sembra, infatti, che il legislatore europeo e quello nazionale dovranno escogitare, in questa materia, dei principi nuovi ed ulteriori, oltre a quello della prevalenza del diritto dell'Unione europea e del principio di cooperazione, per consentire all'interprete di stabilire di volta in volta come si possano svolgere tutte le attività di sorveglianza e deliberative in maniera coordinata.

Il regolamento 2024/1689 prevede, infine, un esteso apparato sanzionatorio. La violazione delle previsioni del regolamento è punita con un sistema di sanzioni pecuniarie commisurate, in misura proporzionale alla tipologia e alla gravità delle violazioni, in una percentuale variabile dal 3% al 7% del fatturato annuo globale dell'impresa autrice dell'illecito. Il regolamento stabilisce massimali più equi per PMI e Start-up al fine di assicurare che i sistemi di IA siano sviluppati ed impiegati in modo etico e legale.

8. *Un approccio timido del legislatore europeo?*

Le regole dell'AI ACT, similmente a quanto accaduto per il GDPR, entreranno in vigore in maniera scaglionata nel tempo, in tal modo consentendo alle imprese e alle pubbliche amministrazioni di adeguarsi progressivamente al nuovo quadro normativo⁶³.

Alla conclusione di tale sommaria disamina del regolamento permangono, tuttavia, diversi dubbi circa la sua reale efficacia e l'idoneità dello stesso a rafforzare un'effettiva sovranità digitale dell'Unione europea. Tale conclusione, infatti, discende da una serie di considerazioni.

In primo luogo, il legislatore europeo non intende affrontare tutti i problemi giuridici posti da detta tecnologia: egli, consapevolmente, mira essenzialmente a disciplinare l'accesso al mercato dei fornitori dei sistemi di intelligenza artificiale. Questa scelta, tuttavia, presenta dei costi ben precisi, in termini di assenza di regolamentazione. Innanzitutto non sono presenti regole uniformi di responsabilità, per i danni provocati dall'uso di sistemi di

A.C. WITT, *Meta v Bundeskartellamt—data-based Conduct between Antitrust Law and Regulation*, in *Journal of Antitrust Enforcement*, 2024, p. 345 ss.; M. PREVIADELLO, *The Duty of Sincere Cooperation between National Competition Authorities and Supervisory Authorities Established by the GDPR: Meta Platforms and Others*, in *Blog-AISDUE*, 2024.

⁶³ Nello specifico, il regolamento è entrato in vigore 20 giorni dopo la pubblicazione nella Gazzetta ufficiale UE; esso si applicherà a far data dal 2 agosto 2026. Tuttavia, mentre i divieti relativi alle pratiche vietate (il c.d. rischio inaccettabile, di cui all'art. 5), si applicano a partire da sei mesi dopo l'entrata in vigore (quindi, già dal 2 febbraio 2025), i codici di buone pratiche, le norme sui sistemi di IA per finalità generali ed il sistema di *governance* si applicheranno solo a partire dal 2 agosto 2026. Infine, gli obblighi per i sistemi ad alto rischio trentasei mesi dopo l'entrata in vigore (2 agosto 2027).

intelligenza artificiale⁶⁴, con la conseguenza che le stesse finiranno per dipendere essenzialmente e dalle regole di diritto internazionale privato contenute nel regolamento c.d. Roma II, sulla legge applicabile alle obbligazioni non contrattuali⁶⁵.

Ma soprattutto, cosa assai più grave, il legislatore europeo ha ommesso di prevedere specifiche misure di protezione dei diritti fondamentali. Infatti, al di là dell'affermazione di principio, secondo cui il legislatore europeo intende proteggere i diritti fondamentali, nel regolamento sull'intelligenza artificiale mancano, in realtà, specifiche forme di tutela per i singoli, con la conseguenza che queste sono eminentemente quelle previste dal regolamento sulla protezione dei dati⁶⁶. Viene dunque da domandarsi se l'AI Act sia effettivamente protettivo per i diritti fondamentali⁶⁷. Tanto più che la valutazione dell'impatto sui diritti fondamentali è, ai sensi dell'art. 26, rimesso agli stessi *deployer*: è agevole pensare che gli utilizzatori del sistema finiranno, nella maggior parte dei casi, con il minimizzare i rischi.

In secondo luogo l'AI Act lascia forse troppo margine agli Stati membri. Molte norme, infatti, non sono *self-executing* e necessitano, pertanto, di ulteriori provvedimenti legislativi da parte degli Stati membri per avere effetto diretto ed essere applicabili.

Un'altra criticità concerne il rischio di mancato coordinamento e le sovrapposizioni, finanche nella formulazione di talune disposizioni, con altri atti normativi adottati dalle istituzioni europee. Il problema si pone, in particolar modo, con il citato Regolamento sulla protezione dei dati personali. Alle volte, la circostanza che determinati diritti siano previsti da entrambe le fonti, come avviene ad esempio con il diritto a non essere sottoposto ad una decisione integralmente automatizzata, potrebbe finanche far sorgere dei dubbi su quale sia la reale portata del diritto che viene concretamente invocato in giudizio.

Inoltre il legislatore ha ommesso di disciplinare aspetti rilevanti nel funzionamento dei sistemi di intelligenza artificiale, quali, in particolare, la tutela dell'ambiente. Questa è sì menzionata in vari considerando⁶⁸, ma di fatto non è affatto regolamentata se non in misura estremamente marginale⁶⁹. In tal modo il regolamento finisce per lasciare scoperto proprio

⁶⁴ È noto, infatti, che il Parlamento europeo aveva a suo tempo emanato una risoluzione diretta alla Commissione europea affinché questa presentasse una proposta su regole armonizzate di responsabilità per l'uso dei sistemi di intelligenza artificiale (Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale, 2020/2014(INI)). Questa, tuttavia, non ha avuto seguito, preferendo la Commissione adottare la proposta che poi ha condotto all'adozione del regolamento 2024/1689. In assenza di una normativa uniforme, sono state sostenute eminentemente due tesi, in entrambi i casi tendenti ad ipotizzare una sorta di regime di responsabilità aggravata in capo al fornitore. Secondo il primo orientamento la responsabilità costituirebbe un regime da "colpa presunta". Secondo un altro orientamento, la responsabilità del fornitore sarebbe assimilabile alla categoria della responsabilità oggettiva, sulla falsariga della responsabilità del produttore nella direttiva da danno da prodotto. In dottrina v., quali esponenti delle due tesi, E. BOCCHINI, *La regolazione giuridica dell'intelligenza artificiale*, Torino, 2024, p. 101 ss.; A. ALBANESE, *La responsabilità extracontrattuale per l'utilizzo dei sistemi di I.A.*, in R. BOCCHINI (a cura di), *Trattato. Le Piattaforme digitali e-Agorà*, Torino, 2025, p. 506 ss., 519.

⁶⁵ Regolamento (CE) n. 864/2007 del Parlamento europeo e del Consiglio, dell'11 luglio 2007, sulla legge applicabile alle obbligazioni extracontrattuali (Roma II), in GUUE L 199 del 31.7.2007, pp. 40–49. In dottrina v. B. CAPIELLO, *AI-systems and Non-Contractual Liability*, Torino, 2022.

⁶⁶ V. G. FINOCCHIARO, *Diritto dell'intelligenza artificiale*, cit., p. 16.

⁶⁷ V. F. PAOLUCCI, *Shortcomings of the AI Act. Evaluating the New Standards to Ensure the Effective Protection of Fundamental Rights*, in *Verfassungsblog*, 14 marzo 2024; E. CIRONE, *L'AI Act e l'obiettivo (mancato?) di promuovere uno standard globale per la tutela dei diritti fondamentali*, in *Quaderni AISDUE*, Fascicolo speciale n. 2/2024, *La nuova disciplina UE sull'intelligenza artificiale*.

⁶⁸ V., ad es., i considerando n. 1, 2, 8, 27, 48, 130, 142, 155, 176.

⁶⁹ V., in particolare, l'art. 40, par. 1, del regolamento: «La richiesta di normazione chiede inoltre prodotti relativi a processi di comunicazione e documentazione *intesi a migliorare le prestazioni dei sistemi di IA in termini di risorse*,

uno degli aspetti negativi maggiormente impattanti dell'uso dei sistemi di intelligenza artificiale: ovvero le conseguenze che questo apporterà sul pianeta. Si tratta di un tema tuttora assai controverso, sia nella letteratura scientifica, che in quella specificamente giuridica; e che è sovente fonte di dibattito e di visioni contrastanti⁷⁰, rispetto al quale, il legislatore europeo – così attento alla tutela dell'ambiente in altri settori normativi (non fosse altro per il disposto dell'art. 11 TFUE) – bene avrebbe potuto introdurre previsioni rilevanti anche in questa materia.

Dal punto di vista economico, poi, occorre chiedersi se questo atto, piuttosto che rafforzare la c.d. sovranità digitale dell'Unione europea, non finisca con il minare la competitività delle imprese europee, le quali sono destinatarie di una serie molto nutrita di obblighi. Parrebbe quindi legittimo porre il dubbio se tale sistema normativo non finisca con il produrre il risultato opposto ovvero quello di aumentare il rischio di delocalizzazione.

L'ultima questione è cercare comprendere se l'AI Act sia in grado di duplicare il successo del GDPR e rappresentare uno standard anche per Stati terzi. Capire, in altri termini, se questo regolamento possa effettivamente produrre un "effetto Bruxelles".

In realtà, la risposta, almeno allo stato, sembrerebbe negativa, perché lo stesso regolamento lascia troppo spazio alla regolamentazione privata, con ciò stesso diminuendone la portata e l'efficacia normativa. Inoltre esso affida un ruolo eccessivamente rilevante alla normativa delegata da parte della Commissione. L'organo esecutivo potrà infatti adottare atti delegati entro cinque anni dall'entrata in vigore (ed anche in futuro, se i colegislatori non revocheranno la delega), per: *i*) escludere che taluni sistemi, che non presentano rischi significativi per la salute ed i diritti fondamentali, siano ad alto rischio (6, par. 6); *ii*) aggiungere o modificare i casi d'uso dei sistemi ad alto rischio (7, par.1); *iii*) stabilire gli obblighi di conservazione della documentazione tecnica per aggiornarla al progresso tecnico; *iv*) adeguare i requisiti di conformità al progresso tecnico (43, parr. 5 e 6); stabilire il contenuto della dichiarazione di conformità (art. 47, par. 5).

In breve, vi sono margini di discrezionalità particolarmente ampi per gli Stati membri; troppa autonomia per i privati; eccessivi poteri normativi, su punti fondamentali della normativa regolamentare, in capo all'esecutivo europeo. Infine, decisamente numerose sono le autorità di *governance*, a vari livelli (europeo e nazionale), chiamate a vigilarne l'applicazione: esse assai difficilmente riusciranno ad operare in maniera fluida e condivisa, soprattutto perché nei rapporti reciproci manca un'autorità sovraordinata, come invece si verifica in materia di concorrenza. Sulla base di questi elementi, sembra necessario concludere che appare assai improbabile che il regolamento possa effettivamente rappresentare uno standard normativo suscettibile di imporsi a livello mondiale⁷¹.

come la riduzione del consumo di energia e di altre risorse del sistema di IA ad alto rischio durante il suo ciclo di vita, e allo sviluppo efficiente sotto il profilo energetico dei modelli di IA per finalità generali» (corsivo aggiunto).

⁷⁰ V. K. CRAWFORD, *Né intelligente, né artificiale. Il lato oscuro dell'IA*, Bologna, 2021, p. 49 ss.; M.A. ALLOGHANI, *Artificial Intelligence and Sustainability*, London, 2023, spec. p. 65 ss.; nonché gli scritti pubblicati nel volume C. UBALDO, K. ALEXANDER (a cura di), *Sostenibilità e intelligenza artificiale: resilienza o panacea?*, Milano, 2025. Per gli aspetti giuridici v. C. DI FRANCESCO MAESA, *Economia circolare e IA: a che condizioni è una sfida possibile per l'UE?*, in *Quaderni AISDUE*, Convegno di Siena 4 e 5 dicembre 2024.

⁷¹ V. anche i dubbi di M. BASSINI, *Oggetto, campo di applicazione e ambito territoriale*, cit., p. 139 s.; U. PAGALLO, *Why the AI Act Won't Trigger a Brussels Effect*, in corso di pubblicazione nel volume *AI Approaches to the Complexity of Legal Systems*, Springer, ma già consultabile al sito https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4696148



FIAMMETTA BORGIA*

IL MODELLO EUROPEO DI REGOLAZIONE DELL'INTELLIGENZA ARTIFICIALE: TRA AMBIZIONE ETICA E IPERTROFIA NORMATIVA

SOMMARIO: 1. Introduzione – 2. Il quadro internazionale in materia di intelligenza artificiale: limiti dell'approccio generalista e prospettive normative. – 3. La Convenzione quadro del Consiglio d'Europa sull'IA: criticità applicative di un approccio regolatorio etico e orizzontale. – 4. Il modello normativo dell'Unione europea sull'IA: tra idealismo regolatorio e vincoli applicativi. – 5. Conclusioni.

1. Introduzione

L'intelligenza artificiale (IA) si configura come una tecnologia capace di operare in modo autonomo, assumendo decisioni sulla base dell'analisi dell'ambiente e orientandosi al conseguimento di obiettivi predeterminati¹. L'evoluzione rapida e pervasiva di tali sistemi, unita alla loro capacità di incidere su ambiti eterogenei e sensibili della vita individuale e collettiva, ha sollevato interrogativi sempre più pressanti circa la necessità di un inquadramento giuridico adeguato. Proprio la natura autonoma, adattiva e, talvolta, opaca dell'IA evidenzia la difficoltà di ricondurre tali tecnologie entro schemi normativi tradizionali, sollecitando gli Stati a interrogarsi su quali strumenti elaborare per affrontare le sfide della cosiddetta "quarta rivoluzione industriale".

La Comunità internazionale, a fronte della natura transnazionale dell'IA, sta tentando di sviluppare strumenti volti a regolamentare la materia in modo coerente e uniforme, con l'obiettivo di garantire un equilibrio tra innovazione tecnologica e tutela dei diritti fondamentali. Tuttavia, gli sforzi compiuti finora, se da un lato evidenziano una crescente consapevolezza circa la necessità di un'azione normativa globale, dall'altro mettono in luce i limiti intrinseci di un approccio olistico e generalista, finora adottato, rivelando alcune criticità significative.

* Professoressa ordinaria di diritto internazionale presso il Dipartimento di giurisprudenza dell'Università degli studi di Roma Tor Vergata

¹ Varie sono le definizioni di IA elaborate recentemente. Si veda, ad esempio, S. LEGG-M. HUTTER, *A Collection of Definitions of Intelligence*, in *Frontiers in Artificial Intelligence and Applications*, 2007, 17-24; R.J. STERNBERG, *The Concept of Intelligence*, in R.J. STERNBERG (ed.), *Handbook of Intelligence*, Cambridge, 2012, 3-15.

In primo luogo, si può osservare che il fine di questi strumenti regolatori sia spesso orientato a limitare l'uso dell'IA, piuttosto che a promuoverne la gestione responsabile. Questo approccio difensivo potrebbe avere come effetto non solo quello di rallentare lo sviluppo tecnologico, ma anche impedire di cogliere appieno le opportunità offerte dall'IA, come strumento utile per affrontare problemi globali, come il cambiamento climatico, le pandemie e la povertà. L'enfasi sull'obiettivo della riduzione dei rischi associati all'IA, infatti, ci sembra conduca a un'ipertrofia regolamentare, spesso eccessivamente restrittiva, che può rallentare o addirittura impedire lo sviluppo di tecnologie potenzialmente rivoluzionarie in settori chiave. Inoltre, l'estrema difficoltà di elaborare strumenti vincolanti e a carattere universale in materia, ci sembra impedisca di affrontare efficacemente le sfide transfrontaliere legate all'IA. Infine, l'approccio olistico adottato da molte iniziative regolatorie non sembra tenere conto della complessità e della diversità delle applicazioni dell'IA nei settori più vari, come, ad esempio, la sanità, l'automazione industriale, la finanza o la sicurezza, che necessiterebbero invece di strumenti *ad hoc*. L'approccio olistico e generalista appare, pertanto, inefficace o, peggio ancora, di ostacolo all'innovazione in alcuni settori, senza che sia garantita un'adequata protezione in altri.

Alla luce di queste considerazioni, l'obiettivo di questo scritto è duplice. In primo luogo, si intende sottolineare la necessità di elaborare un approccio giuridico diverso alla questione, operando lo sforzo di spostare l'obiettivo regolatorio dalla limitazione all'utilizzo responsabile dell'IA. Il fine degli strumenti giuridici internazionali ci sembra, infatti, che debba essere quello di creare un ambiente favorevole per lo sviluppo e l'uso etico dell'IA, piuttosto che quello di limitarne i rischi. Ciò evidentemente richiede un maggiore impegno nella promozione della trasparenza, nella definizione di responsabilità chiare e nella creazione di meccanismi di monitoraggio e cooperazione. In secondo luogo, si vuole sottolineare la necessità di superare l'approccio olistico sino ad ora prevalente per privilegiare una regolamentazione settoriale. Un quadro settoriale ci sembra, infatti, possa essere più idoneo ad affrontare le specificità delle diverse applicazioni dell'IA, garantendo una maggiore effettività normativa.

A tal fine, l'analisi che segue si concentrerà sul cosiddetto “modello europeo” che, pur distinguendosi per l'ambizione etica, appare privo di una visione sufficientemente orientata alla concreta applicazione. La Convenzione quadro del Consiglio d'Europa sull'intelligenza artificiale e il Regolamento dell'Unione europea sull'IA rappresentano, infatti, espressioni emblematiche della marcata ipertrofia regolamentare che caratterizza l'approccio europeo, impegnato in una complessa e faticosa ricerca di equilibrio tra progresso tecnologico e tutela dei diritti umani, più che nella definizione di un quadro normativo funzionale e concretamente applicabile. Proprio l'esame di questi strumenti mostrerà, dunque, i limiti sistemici dell'approccio olistico europeo, sinora valorizzato, evidenziando la necessità di sviluppare piuttosto iniziative regolatorie settoriali, anche superando la “retorica” del rischio.

2. Il quadro internazionale in materia di intelligenza artificiale: limiti dell'approccio generalista e prospettive normative

Nel contesto del diritto internazionale, il principale obiettivo degli Stati in tema di IA sembra essere sino ad ora quello di “limitarne e quindi garantirne” un utilizzo responsabile, in particolare per quanto riguarda quella generativa. Tale tendenza riflette una logica di

contenimento più che di promozione di uno sviluppo normativamente accompagnato e responsabile. Nonostante il carattere frammentario degli interventi in materia, emergono tentativi di elaborare una risposta giuridica sistemica alle sfide poste dall'interazione tra sviluppo tecnologico, strutture democratiche e principi etici

Il dibattito in dottrina, seppur embrionale, è estremamente vitale². Nel tentativo non semplice di individuare alcune direttrici giuridiche per il fenomeno in esame, è possibile delineare, seppur con cautela, un quadro normativo strutturato su diversi livelli, che comprende non solo le fonti internazionali in senso stretto, ma anche gli strumenti di *soft law*, i provvedimenti nazionali con impatti transnazionali, le norme tecniche e le misure di autoregolamentazione delle imprese.

In questa prospettiva, andrebbero innanzitutto considerate le numerose norme di diritto internazionale già esistenti che possono essere applicate anche in materia di IA, come quelle a protezione dei diritti umani fondamentali³, o relative al diritto internazionale umanitario⁴, nonché alcune convenzioni più recenti, specificamente rivolte a fenomeni legati allo sviluppo tecnologico digitale, ad esempio nell'ambito della protezione dei dati o la Convenzione 108+ del Consiglio d'Europa⁵.

Inoltre, la materia è al centro di numerose iniziative da parte di alcune organizzazioni intergovernative, che si sforzano di configurare un quadro giuridico *ad hoc* per lo spazio digitale⁶. In particolare, dai lavori delle Nazioni Unite emerge l'esigenza di applicare il diritto internazionale alle sfide della digitalizzazione, inclusa l'IA, soprattutto ai fini del mantenimento della pace e della sicurezza internazionale⁷. Tuttavia, tali iniziative, sebbene

² Sul punto si veda, tra i più recenti studi, T. BURRI, *International Law and Artificial Intelligence*, in *German Yearbook of International Law*, 2017, 91-108; E. BENVENISTI, *Upholding Democracy Amid the Challenges of New Technology: What Role for the Law of Global Governance?*, in *European Journal of International Law*, 2018, 9-82; M. LILJEFORS-G. NOLL-D. STEUER, *War and Algorithm*, London, 2019; M.M. MAAS, *International Law Does Not Compute: Artificial Intelligence and the Development, Displacement or Destruction of the Global Legal Order*, in *Melbourne Journal of International Law*, 2019, 29-57; P. MOLNAR, *Technology on the Margins: AI and Global Migration Management from a Human Rights Perspective*, in *Cambridge International Law Journal*, 2019, 305-330; A. DEEKS, *Introduction to the Symposium: How Will Artificial Intelligence Affect International Law?*, in *American Journal of International Law Unbound*, 2020, 138-140; M. LANGFORD, *Taming the Digital Leviathan: Automated Decision-Making and International Human Rights*, in *American Journal of International Law Unbound*, 2020, 141-146; M. HILDEBRANDT, *Text-Driven Jurisdiction in Cyberspace*, in *Theoretical and Applied Law*, 2021, 6-20.; A. HÁRS, *AI and International Law: Legal Personality and Avenues for Regulation*, in *Hungarian Journal of Legal Studies*, 2022, 320-344; M. ARVIDSSON-G. NOLL, *Artificial Intelligence, Decision Making and International Law*, in *Nordic Journal of International Law*, 2023, 1-8.

³ In relazione alle possibili implicazioni dei sistemi di IA sui diritti umani si fa riferimento soprattutto al diritto alla *privacy*, alla libertà di espressione ed al divieto di discriminazione, enunciati – come è noto – nelle disposizioni della Dichiarazione universale dei diritti dell'uomo, della Convenzione europea dei diritti dell'uomo e del Patto sui diritti civili e politici. Si veda sul punto A. BEDUSCHI, *Human Rights and the Governance of Artificial Intelligence*, Geneva Academy of International Humanitarian Law and Human Rights, 2020, reperibile online.

⁴ H. NASU, *Artificial Intelligence and the Obligation to Respect and to Ensure Respect for IHL*, in E. MASSINGHAM-A. MCCONNACHIE (eds.), *Ensuring Respect for International Humanitarian Law*, I ed., London, 2020.

⁵ Al testo originario della Convenzione si affianca il Protocollo n. 223, noto come Convenzione 108+, non ancora entrato in vigore, che comprende nuovi diritti delle persone riguardo a processi decisionali basati su algoritmi, che assumono particolare rilevanza nell'ambito dello sviluppo dell'intelligenza artificiale.

⁶ S. FU-A. XING, *International Soft Law Governance of Artificial Intelligence: Advantages, Approaches, and Credibility*, in *Journal of Business Theory and Practice*, 2023, 35-45.

⁷ Sul dibattito in seno alle Nazioni Unite sull'intelligenza artificiale v. K.J. SMITH, *The AI Community and the United Nations: A Missing Global Conversation & A Closer Look at Social Good*, in *Association for Advancement of Artificial Intelligence (AAAI) Spring Symposium Series*, 2017, reperibile online; E. FOURNIER-TOMBS, *Towards a United Nations Internal Regulation for Artificial Intelligence*, in *Big Data & Society*, 2021, 1 ss.

importanti sul piano dei principi, presentano spesso un contenuto generico e non vincolante, con limitate ricadute operative.

Senza, dunque, avere alcuna pretesa esaustiva, in questa sede ci sembra opportuno almeno richiamare alcune tra le più importanti proposte in tal senso.

Nel sistema delle Nazioni Unite, nel 2021, l'UNESCO ha pubblicato le linee guida sull'uso etico dell'intelligenza artificiale⁸. Queste evidenziano quali principi di riferimento l'equità, la trasparenza, la tutela della *privacy*, la responsabilità e lo sviluppo sostenibile. Nello stesso anno la Conferenza delle Nazioni Unite sul commercio e lo sviluppo (UNCTAD) ha pubblicato una serie di principi sull'intelligenza artificiale e la *governance* dei dati affrontando questioni chiave come il flusso dei dati⁹. Il Consiglio per i diritti umani, nel rapporto del 2021 sull'intelligenza artificiale, ha proposto una serie di principi e linee guida per garantire che le nuove tecnologie basate sull'IA siano sviluppate e applicate in linea con gli *standard* sui diritti umani¹⁰.

Anche l'Organizzazione per la cooperazione e lo sviluppo economico (OCSE), nel 2019, ha pubblicato una serie di principi di carattere sociale, giuridico ed etico sull'intelligenza artificiale¹¹, nei quali si afferma, da un lato, che lo sviluppo e l'applicazione delle nuove tecnologie basate sull'IA dovranno perseguire i principi di giustizia e uguaglianza evitando discriminazioni nei confronti di determinati gruppi o individui, dall'altro, che nei sistemi di IA dovranno essere rispettate le norme relative alla protezione dei dati e dovranno essere adottate misure per garantire la sicurezza e la riservatezza di questi ultimi per prevenire abusi e fughe di dati.

In ambito europeo, il Consiglio d'Europa, nel 2019, su mandato del Comitato dei Ministri, ha istituito un gruppo consultivo *ad hoc* per l'IA al fine di elaborare un accordo in materia di sviluppo, progettazione e utilizzo dell'IA. Nel 2020, il Comitato ha proposto nove principi base per i futuri strumenti giuridici in materia, tra i quali troviamo citati i principi di non-discriminazione, di trasparenza dei sistemi di IA, di salvaguardia della dignità umana, nonché i diritti alla protezione dei dati e alla *privacy*¹². Più recentemente, nel 2024, sotto l'egida dell'organizzazione è stato concluso il primo trattato sull'IA. La Convenzione quadro del Consiglio d'Europa sull'intelligenza artificiale e i diritti umani, la democrazia e lo Stato di diritto¹³ è stata conclusa il 17 maggio 2024 a Strasburgo durante l'incontro ministeriale annuale del Comitato dei Ministri del Consiglio d'Europa. Aperta all'adesione anche da parte di Stati non europei, la Convenzione istituisce un quadro normativo organico applicabile all'intero ciclo di vita dei sistemi di IA, affrontando i potenziali rischi connessi al loro utilizzo e promuovendo, al contempo, un'innovazione tecnologica responsabile. L'accordo adotta

⁸ UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, adopted on 23 November 2021, SHS/BIO/PI/2021/1, reperibile *online*.

⁹ UNCTAD, *Technology and Innovation Report 2021*, pubblicato il 25 febbraio 2021, UNCTAD/TIR/2020, reperibile *online*.

¹⁰ Human Rights Council, *The right to privacy in the digital age: Report of the United Nations High Commissioner for Human Rights*, 13 September 2021, A/HRC/48/31, reperibile *online*.

¹¹ OCSE, *Recommendation of the Council on Artificial Intelligence*, adopted on 22 May 2019, OECD/LEGAL/0449 (amended on 03 May 2024).

¹² Ad Hoc Committee On Artificial Intelligence (CAHAI), *Towards Regulation of AI Systems*, December 2020, COE Study DGI (2020)16, reperibile *online*.

¹³ Cfr. Committee on Artificial Intelligence (CAI), *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, 17 May 2024, CM(2024)52-final. Tale Convenzione, al momento della pubblicazione di questo scritto, è stata firmata (ma ancora non ratificata) da Andorra, Georgia, Islanda, Montenegro, Norvegia, Moldavia, San Marino, Regno Unito, nonché da Israele, Stati Uniti e dall'Unione europea.

appunto un approccio olistico basato sulla valutazione dei rischi e impone un esame approfondito delle possibili conseguenze negative derivanti dalla progettazione, dallo sviluppo, dall'impiego e dalla dismissione dei sistemi di IA, con il fine di garantire che l'evoluzione tecnologica sia compatibile con la tutela dei diritti fondamentali e con il rispetto dei principi democratici e dello Stato di diritto. Tuttavia, la scelta di un approccio marcatamente olistico e basato sul rischio ne limita fortemente l'efficacia applicativa.

Infine, per quanto riguarda l'Unione europea, nel dicembre 2023, i membri del Consiglio dell'Unione europea e i delegati del Parlamento europeo hanno raggiunto un accordo sulla necessità di creare regole armonizzate sull'IA¹⁴. Il cosiddetto "Regolamento sull'intelligenza artificiale" ha, infatti, l'obiettivo di assicurare che i sistemi di IA immessi sul mercato europeo e utilizzati nell'UE siano sicuri e rispettino i diritti fondamentali e i valori dell'UE. Si tratta, anche in questo caso, di un'iniziativa legislativa finalizzata a guidare lo sviluppo e l'adozione di un'IA da concepire come "sicura e affidabile", sia da parte di enti pubblici che privati nell'ambito del mercato unico europeo. Il regolamento si basa su un approccio "basato sul rischio", dove la severità delle regole aumenta in relazione al potenziale danno che l'IA può arrecare alla società. Quest'ultima, costituisce – come è noto – la prima iniziativa legislativa di questo genere a livello mondiale e potrebbe stabilire uno *standard* internazionale per la regolamentazione dell'IA in altre regioni, similmente a quanto avvenuto con il Regolamento generale sulla protezione dei dati. Anche in questo caso, però, l'adozione di un modello generalista ed extraterritoriale presenta importanti limiti di effettività.

Quanto agli strumenti di *soft law*, ai provvedimenti nazionali con impatti transnazionali, alle norme tecniche e alle misure di autoregolamentazione delle imprese, che ad oggi completano il quadro internazionale in materia, si tratta soprattutto di strumenti giuridicamente non vincolanti, che concorrono a fornire una guida per la comunità internazionale per promuovere lo sviluppo giuridico ed etico nel campo dell'intelligenza artificiale. Anche in questo caso, gli strumenti richiamati appaiono accomunati da una visione comune che propende per uno sviluppo di un impianto sistemico relativo all'intelligenza artificiale che sia quanto più possibile antropocentrico e riflettono altresì una preoccupazione generalizzata circa i potenziali effetti negativi che l'utilizzo di tali tecnologie potrebbe produrre sulla tutela dei diritti fondamentali¹⁵. Manca, invece, un'effettiva articolazione operativa che consenta di passare dalla teoria alla prassi.

Da questa sommaria ricognizione dei principali strumenti giuridici esistenti emerge, dunque, come sia a livello universale che nell'ambito di quadri giuridici sovranazionale la normativa in materia di intelligenza artificiale si stia via via delineando secondo un binario ben definito, orientato a limitare più che gestire questa tecnologia (in chiave antropocentrica) e ad offrire medesime (o analoghe) direttrici giuridiche per qualsiasi ambito applicativo dell'IA.

La principale debolezza di questo approccio risiede nell'asimmetria tra l'enfasi posta sulla limitazione preventiva e l'elaborazione di strumenti normativi flessibili e applicabili, capaci di adattarsi ai ritmi dell'evoluzione tecnologica. Sul piano giuridico, infatti, l'approccio

¹⁴ Cfr. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance), 12.7.2024, EN L series 2024/1689.

¹⁵ Si veda in tal senso 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, 782-810, in particolare p. 792.

antropocentrico, sebbene fondato su valori fondamentali, tende a cristallizzare una visione difensiva, non considerando adeguatamente la complessità delle interazioni tra IA e sistemi socio-economici globali. Infine, l'idea di fornire soluzioni normative generali per ogni attività svolta grazie all'utilizzo dell'IA risulta poco realistica. Una tale regolamentazione non riesce a cogliere le peculiarità dei diversi settori di applicazione, compromettendo la capacità di rispondere in modo mirato e proporzionato ai rischi concreti posti dalla tecnologia. Questo limita la possibilità di costruire un quadro giuridico efficace e dinamico che possa contemperare tutela dei diritti e innovazione tecnologica.

Tali criticità si riflettono chiaramente anche nei più recenti strumenti normativi, come la Convenzione quadro del Consiglio d'Europa sull'intelligenza artificiale e il Regolamento dell'Unione europea sull'intelligenza artificiale. Pur prevedendo una qualche forma di applicazione extraterritoriale – volta, in teoria, a incidere anche al di fuori dei confini strettamente europei – tali strumenti appaiono ancora inadeguati a perseguire concretamente gli ambiziosi obiettivi che si propongono. L'approccio regolatorio si caratterizza infatti per un'eccessiva rigidità e per un'impostazione ipertrofica, poco compatibile con la dinamicità e l'eterogeneità dei contesti globali in cui l'IA si sviluppa. Ne risulta una governance frammentata e poco efficace, incapace di assicurare un reale bilanciamento tra tutela dei diritti fondamentali e promozione dell'innovazione tecnologica su scala globale.

3. *La Convenzione quadro del Consiglio d'Europa sull'IA: criticità applicative di un approccio regolatorio etico e orizzontale*

La Convenzione quadro del Consiglio d'Europa sull'IA rappresenta uno strumento normativo di notevole rilevanza per il diritto internazionale contemporaneo, non solo per la sua ambizione di regolamentare un settore in rapida evoluzione, ma anche per il suo tentativo di affrontare, con soluzioni vincolanti per gli Stati contraenti, il delicato equilibrio tra innovazione tecnologica e tutela dei diritti fondamentali.

L'obiettivo, infatti, è quello di regolamentare l'intero ciclo di vita dei sistemi di IA¹⁶, al fine di garantire la tutela della dignità umana, della democrazia e dello Stato di diritto, come stabilito dall'articolo 3 della Convenzione, e di offrire un bilanciamento tra interessi apparentemente contrapposti. Tuttavia, al di là delle dichiarazioni di intenti, ci sembra – sin da una prima lettura – che si tratti più di una enunciazione di principi che di uno strumento operativo idoneo a fronteggiare le sfide concrete poste dall'IA. Appaiono numerosi, infatti, i dubbi sulla sua effettiva capacità di coniugare progresso tecnologico e tutela dei diritti fondamentali, mentre ci sembra più concreto il rischio di rappresentare un quadro normativo formale che abbia scarsa capacità di essere applicato nella realtà.

Già nel preambolo della Convenzione appare chiaro il suo limite principale: i potenziali benefici delle tecnologie di IA sono appena accennati, mentre i rischi legati al suo utilizzo

¹⁶ Il concetto di “ciclo di vita” (life cycle) dell'intelligenza artificiale include una serie di passaggi, che vanno dalla fase iniziale di preparazione e programmazione, fino alla sua vendita ed eventuale dismissione: «pre-design, design, development, evaluation, testing, deployment, use, sale, procurement, operation and decommissioning». Nel medesimo senso si erano già espresse le Nazioni Unite. Si veda in proposito Risoluzione dell'Assemblea Generale delle Nazioni Unite del 21 marzo 2024, *Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development*, A/RES/78/265, reperibile online.

ispirano l'intera regolamentazione¹⁷. Infatti, sebbene nel testo siano riconosciute formalmente le opportunità offerte dall'IA per promuovere prosperità e innovazione, le disposizioni successive finiscono per privilegiare una retorica del rischio, imponendo alcuni obblighi stringenti, che possono avere impatti negativi ingiustificati sul suo sviluppo tecnologico¹⁸. Tale approccio emerge chiaramente nelle misure che vietano l'utilizzo di determinate categorie di IA, quali i sistemi di profilazione, di valutazione sociale, di algoritmi decisionali automatizzati, di riconoscimento biometrico in tempo reale e sistemi predittivi utilizzati per la sicurezza pubblica, ritenute ad alto rischio per i diritti fondamentali e la democrazia, senza prevedere meccanismi adeguati a valutare i benefici che esse potrebbero offrire in altri settori, come la salute, l'istruzione o la sostenibilità ambientale.

Se la visione ispiratrice dello strumento in esame appare sbilanciata, anche l'ambito di applicazione prescelto pone alcuni problemi. La Convenzione, infatti, si riferisce a tutto il ciclo di vita delle intelligenze artificiali che possono influenzare i diritti fondamentali, la democrazia e lo Stato di diritto, rimanendo esclusi i soli sistemi di IA utilizzati per la tutela degli interessi relativi alla sicurezza nazionale, per le attività di ricerca volte a testare i sistemi di IA e per le materie relative alla difesa nazionale¹⁹. Questa scelta appare contraddittoria rispetto all'ambizione di creare un quadro normativo globale e coerente. Sebbene sia comprensibile la necessità di salvaguardare la sovranità degli Stati membri in settori sensibili, tale esclusione riduce l'efficacia complessiva della Convenzione, lasciando spazio a potenziali lacune normative che potrebbero essere sfruttate dagli Stati stessi o dagli operatori privati per aggirare gli obblighi previsti.

Inoltre, l'approccio adottato dalla Convenzione risente delle difficoltà di armonizzare le esigenze e le sensibilità di quarantasette Stati membri con tradizioni giuridiche differenti. Sebbene sia stato compiuto uno sforzo per stabilire principi universali, come trasparenza,

¹⁷ Si veda in proposito *Council of Europe Framework Convention on Artificial Intelligence and Human Rights*, cit., Preambolo, che dopo aver riconosciuto «that activities within the lifecycle of artificial intelligence systems may offer unprecedented opportunities to protect and promote human rights, democracy and the rule of law», afferma: «Concerned that certain activities within the lifecycle of artificial intelligence systems may undermine human dignity and individual autonomy, human rights, democracy and the rule of law; Concerned about the risks of discrimination in digital contexts [...]; Concerned by the misuse of artificial intelligence systems and opposing the use of such systems for repressive purposes [...]».

¹⁸ *Ibidem*, art. 8 – Transparency and oversight: «Each Party shall adopt or maintain measures to ensure that adequate transparency and oversight requirements tailored to the specific contexts and risks are in place in respect of activities within the lifecycle of artificial intelligence systems, including with regard to the identification of content generated by artificial intelligence systems»; art. 16 – Risk and impact management framework: «Each Party shall, taking into account the principles set forth in Chapter III, adopt or maintain measures for the identification, assessment, prevention and mitigation of risks posed by artificial intelligence systems by considering actual and potential impacts to human rights, democracy and the rule of law. Such measures shall be graduated and differentiated, as appropriate, and: a) take due account of the context and intended use of artificial intelligence systems, in particular as concerns risks to human rights, democracy, and the rule of law; b) take due account of the severity and probability of potential impacts; c) consider, where appropriate, the perspectives of relevant stakeholders, in particular persons whose rights may be impacted; d) apply iteratively throughout the activities within the lifecycle of the artificial intelligence system; e) include monitoring for risks and adverse impacts to human rights, democracy, and the rule of law; f) include documentation of risks, actual and potential impacts, and the risk management approach; and g) require, where appropriate, testing of artificial intelligence systems before making them available for first use and when they are significantly modified [...]».

¹⁹ *Ibidem*, art. 3, parr. 2, 3, e 4.

responsabilità, sostenibilità e innovazione sicura²⁰, la loro realizzazione concreta risulta complessa e frammentata. Ne deriva un quadro normativo inefficace nel rispondere alle sfide globali poste dall'IA, quali, ad esempio, la proliferazione di contenuti manipolati o il rafforzamento delle disuguaglianze sociali. Peraltro, la mancanza di disposizioni specifiche per incentivare lo sviluppo di tecnologie inclusive e sostenibili relegano questi aspetti a dichiarazioni programmatiche, privandoli di immediata efficacia operativa.

Infine, sempre in termini generali, secondo la Convenzione dalla natura intrinsecamente globale dell'intelligenza artificiale deriverebbe la necessità di un approccio multilaterale e collaborativo²¹. Il testo dell'accordo incoraggia gli Stati firmatari non solo a creare normative nazionali coerenti con i principi della Convenzione, ma anche

a promuovere una cooperazione internazionale che faciliti lo scambio di informazioni, esperienze e buone pratiche. La costruzione di una cooperazione tra Stati in materia sarebbe, infatti, fondamentale dal momento che le applicazioni di IA superano i confini nazionali e sollevano problemi di giurisdizione e responsabilità transnazionali. In altri termini, ciò che si vuole proporre è un paradigma di *governance* multilivello, che tenti di coniugare l'autonomia degli Stati con l'esigenza di un approccio coordinato a livello internazionale. Tale impostazione, seppure condivisibile sul piano teorico, presenta criticità evidenti. Basti pensare alla complessità di armonizzare le normative degli Stati firmatari, che produrrà inevitabilmente compromessi applicativi, senza che siano resi effettivi i principi enunciati nella Convenzione. Inoltre, la cooperazione internazionale auspicata, priva di strumenti vincolanti per favorire lo scambio di dati e buone pratiche, ha un'alta probabilità di restare lettera morta, specialmente nell'attuale contesto geopolitico altamente competitivo. Infine, il paradigma di governance multilivello, pur apprezzabile, soffre dell'assenza di soluzioni chiare per affrontare questioni come quelle relative alla giurisdizione e ai regimi di responsabilità. La mancanza di meccanismi concreti di *follow-up* che facilitino la collaborazione tra Stati e l'assenza di sanzioni per la non conformità rende in definitiva il modello proposto più un ideale teorico che un meccanismo operativo efficace.

Quanto al problema del bilanciamento tra innovazione tecnologica e diritti umani, in concreto la Convenzione introduce il concetto di "proporzionalità del rischio"²²: un principio giuridico che impone di calibrare le misure regolatorie in base alla gravità e alla probabilità dei rischi associati a specifiche applicazioni di IA²³. Inoltre, è prevista la possibilità, per gli Stati che ne faranno parte, di vietare determinate tipologie di intelligenza artificiale che

²⁰ *Ibidem*, art. 8 – Transparency and oversight; art. 9 – Accountability and responsibility; art. 12 – Reliability; art. 13 – Safe innovation.

²¹ *Ibidem*, art. 25: «The Parties shall co-operate in the realisation of the purpose of this Convention. Parties are further encouraged, as appropriate, to assist States that are not Parties to this Convention in acting consistently with the terms of this Convention and becoming a Party to it [...]».

²² *Ibidem*, art. 16. Vedi anche C. NOVELLI-F. CASOLARI-A. ROTOLO, et al., *AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act*, in *Digital Society*, 3, 2024. Un approccio analogo è previsto dal Regolamento dell'Unione europea sull'IA.

²³ Tale rischio, tuttavia, a differenza di quanto avviene nell'ambito del diritto dell'Unione europea (dove la gestione del rischio è attuata in via generale ma poi è ulteriormente specificata con la valutazione d'impatto sui diritti fondamentali), si concentra soprattutto sui rischi che l'IA pone per la democrazia, lo Stato di diritto ed i diritti umani [*Council of Europe Framework Convention on Artificial Intelligence and Human Rights*, cit., art. 16, par. 2, lett. a)]. È dunque ragionevole pensare che la gestione del rischio e la valutazione dell'impatto dell'IA su democrazia, Stato di diritto e diritti fondamentali dovrà coordinarsi, per gli Stati membri dell'Unione europea, con quella prevista dall'*AI Act* [*Regulation (EU) 2024/1689 Of The European Parliament and of the Council of 13 June 2024*, cit., art. 9: Risk management system]. Così in G. ZACCARONI, *Intelligenza artificiale e principio democratico: riflessioni a margine dell'emersione di un quadro normativo europeo*, in *Quaderni AISDUE*, 2, 2024, 1-37, in part. 18.

costituiscano, sulla base degli strumenti di gestione del rischio, un pericolo per la democrazia, i diritti fondamentali e lo Stato di diritto²⁴. Questo approccio non solo garantirebbe un'efficace protezione dei diritti fondamentali, ma soffocherebbe l'innovazione tecnologica attraverso regolamentazioni eccessivamente rigide o generiche. Ci sembra, tuttavia che, sebbene il principio sia condivisibile sul piano teorico, la sua applicazione risulti complessa e problematica. Innanzitutto, la valutazione della gravità e della probabilità dei rischi associati alle applicazioni di IA si basa su criteri spesso vaghi e difficili da standardizzare, data la diversità dei contesti tecnologici e sociali in cui queste applicazioni operano. La mancanza di criteri chiari e univoci per identificare e misurare i rischi può portare a interpretazioni soggettive e incoerenti da parte degli Stati contraenti. Inoltre, la calibrazione delle misure regolatorie in base al rischio presuppone un livello elevato di conoscenza tecnica e giuridica da parte delle autorità competenti, il che potrebbe non essere garantito in tutti i contesti nazionali, finendo per generare un'interpretazione e applicazione frammentata e forse non sempre coerente del principio²⁵. In definitiva, tale principio rischia di trasformarsi in un'astrazione teorica difficilmente applicabile: l'assenza di un quadro operativo dettagliato e l'elevato margine di discrezionalità lasciato agli Stati rischiano di compromettere seriamente l'efficacia della Convenzione nel garantire il fine del raggiungimento di un bilanciamento tra innovazione tecnologica e tutela dei diritti fondamentali.

Un altro aspetto rilevante della Convenzione è l'enfasi posta sulla trasparenza e sulla intellegibilità degli algoritmi²⁶. In un contesto in cui le decisioni automatizzate possono avere un impatto significativo sui diritti individuali, appare, infatti, essenziale che tali decisioni siano comprensibili non solo per gli esperti tecnici, ma anche per gli utenti finali e le autorità di controllo²⁷. La Convenzione richiede che gli sviluppatori di IA forniscano una documentazione adeguata e accessibile che descriva il funzionamento dei sistemi, le logiche sottostanti e i potenziali margini di errore. Tale obbligo sarebbe volto a promuovere la responsabilità degli sviluppatori ed a rafforzare la fiducia pubblica nelle tecnologie di IA, un elemento fondamentale per il loro sviluppo e la loro accettazione su larga scala. Anche in tal caso, tuttavia, l'enfasi sulla trasparenza e intellegibilità degli algoritmi, pur fondamentale, presenta limiti significativi. La complessità tecnica di molti modelli di IA, come quelli basati su reti neurali profonde, rende particolarmente complesso offrire una visione pienamente comprensibile delle loro logiche, persino agli esperti. Inoltre, la mancanza di una definizione

²⁴ *Council of Europe Framework Convention on Artificial Intelligence and Human Rights*, cit., art. 16, par. 4.

²⁵ La necessità di bilanciare efficacemente i diritti fondamentali e l'innovazione tecnologica nei settori sensibili, come il sistema giudiziario o la sorveglianza pubblica, appare evidente; è altrettanto chiara la difficoltà di applicare tali principi in questi ambiti. Sebbene l'introduzione di regolamentazioni più stringenti in questi settori possa essere giustificata dalla necessità di tutela dei diritti umani, l'assenza di strumenti concreti per mitigare gli effetti delle restrizioni sullo sviluppo tecnologico potrebbe rischiare di disincentivare gli investimenti in soluzioni innovative, finendo per spingere le imprese tecnologiche a spostarsi verso contesti normativi meno rigidi e riducendo così l'efficacia concreta della Convenzione e la competitività tecnologica degli Stati firmatari.

²⁶ *Council of Europe Framework Convention on Artificial Intelligence and Human Rights*, cit., art. 8. I. ERDOGAN, *Diving into the Iceberg: Establishing Transparency in AI for Law Enforcement*, in *European Papers*, 9, 2024, 956-977.

²⁷ *Council of Europe Framework Convention on Artificial Intelligence and Human Rights*, cit., art. 14, par. 2, lett. a) e b): «[...] each Party shall adopt or maintain measures including: a) measures to ensure that relevant information regarding artificial intelligence systems which have the potential to significantly affect human rights and their relevant usage is documented, provided to bodies authorized to access that information and, where appropriate and applicable, made available or communicated to affected persons; b) measures to ensure that the information referred to in subparagraph a) is sufficient for the affected persons to contest the decision(s) made or substantially informed by the use of the system, and, where relevant and appropriate, the use of the system itself [...]».

chiara dell'espressione "documentazione adeguata", rischia di generare interpretazioni divergenti tra gli Stati, compromettendo l'efficacia delle norme. Infine, l'obbligo di trasparenza impone costi elevati agli sviluppatori, penalizzando in particolare le piccole imprese, e non garantisce automaticamente maggiore fiducia pubblica: la disponibilità di informazioni tecniche non si traduce necessariamente in una loro comprensione da parte di utenti o autorità. Senza *standard* chiari e programmi di sensibilizzazione, la trasparenza rischia di rimanere un principio formale, incapace di promuovere una reale responsabilità.

Appare inoltre degno di nota il tentativo della Convenzione di introdurre il principio dell'innovazione sicura, che consiste nell'utilizzo di ambienti controllati, detti *sandboxes*, per sperimentare e valutare tecnologie di IA in modo circoscritto, prima della loro implementazione su larga scala²⁸. Questi ambienti consentono di simulare condizioni reali, identificare potenziali rischi e testare soluzioni in modo sicuro, limitando l'impatto di eventuali problemi sull'ambiente esterno. Al di là dell'obiettivo virtuoso, tuttavia, per garantire l'efficacia di tale approccio, ci sembra sia necessario sviluppare un quadro giuridico e operativo più dettagliato, che includa linee guida pratiche sull'uso delle *sandboxes*, criteri chiari per la valutazione dei risultati e modalità di condivisione delle informazioni ottenute tra gli Stati contraenti. Ad esempio, occorrerebbe definire *standard* uniformi per progettare e condurre i *test*, così da assicurare risultati confrontabili e applicabili in diversi contesti. Inoltre, il principio dell'innovazione sicura dovrebbe essere strettamente collegato a iniziative di cooperazione internazionale. Gli Stati membri potrebbero condividere le conoscenze acquisite nelle *sandboxes*, evitando così duplicazioni di sforzi e frammentazioni normative. Questo approccio collaborativo non solo promuoverebbe uno sviluppo tecnologico più armonioso, ma favorirebbe anche l'adozione di *best practices* comuni, aumentando la fiducia nell'IA e accelerando la realizzazione di soluzioni sicure e innovative a livello globale.

La tutela della privacy e della protezione dei dati personali, riconosciuta come pilastro della Convenzione²⁹, si pone in continuità con strumenti come il GDPR definito a livello di UE, introducendo però obblighi specifici per le tecnologie di IA, tra cui le valutazioni d'impatto obbligatorie per i sistemi ad alto rischio³⁰. Questo approccio mira a rafforzare l'*accountability* e a garantire che le violazioni dei diritti siano adeguatamente sanzionate. Tuttavia, anche la previsione delle valutazioni d'impatto, sebbene ambiziosa, soffre di vaghezza operativa, proprio legata alla mancanza di una definizione chiara e uniforme di "sistemi ad alto rischio". Un ulteriore limite è l'onere gravoso imposto agli sviluppatori, che potrebbe ostacolare l'innovazione tecnologica, soprattutto per le piccole e medie imprese. Ciò potrebbe incentivare le aziende a spostarsi verso giurisdizioni con regolamentazioni

²⁸ *Ibidem*, art. 13: «With a view to fostering innovation while avoiding adverse impacts on human rights, democracy and the rule of law, each Party is called upon to enable, as appropriate, the establishment of controlled environments for developing, experimenting and testing artificial intelligence systems under the supervision of its competent authorities». Sulle *sandboxes* (anche se con riferimento all'*AI Act*), vedi G. RUGANI, *La promozione di strumenti di co-regolazione dell'intelligenza artificiale nell'AI Act, con particolare riferimento alle regulatory sandboxes*, in *Quaderni AISDUE*, 2, 2024, 1-15.

²⁹ *Council of Europe Framework Convention on Artificial Intelligence and Human Rights*, cit., art. 11: « Each Party shall adopt or maintain measures to ensure that, with regard to activities within the lifecycle of artificial intelligence systems: a) privacy rights of individuals and their personal data are protected, including through applicable domestic and international laws, standards and frameworks; and b) effective guarantees and safeguards have been put in place for individuals, in accordance with applicable domestic and international legal obligations».

³⁰ *Ibidem*, art. 12: «Each Party shall take, as appropriate, measures to promote the reliability of artificial intelligence systems and trust in their outputs, which could include requirements related to adequate quality and security throughout the lifecycle of artificial intelligence systems».

meno rigide, vanificando l'obiettivo di un'applicazione uniforme delle norme in esame³¹. La Convenzione, infine, non affronta adeguatamente il ruolo cruciale della formazione e della sensibilizzazione³², elementi indispensabili per una corretta gestione delle tecnologie di IA e per garantire un'effettiva protezione dei dati personali.

La regolamentazione della responsabilità giuridica per i danni causati dai sistemi di IA è, senza dubbio, una delle sfide più complesse per il diritto internazionale. Lo strumento convenzionale in esame dedica uno spazio alla responsabilità giuridica per i danni causati dai sistemi di IA³³. Sebbene il testo non fornisca una soluzione definitiva, stabilisce principi generali che possono guidare gli Stati nell'elaborazione di normative nazionali coerenti³⁴, quali il principio di responsabilità oggettiva per le applicazioni ad alto rischio³⁵ e l'obbligo di adottare misure preventive per ridurre al minimo i rischi³⁶. Questi principi rappresentano un punto di partenza fondamentale per affrontare le complesse questioni di attribuzione della responsabilità in un contesto in cui le decisioni automatizzate possono essere il risultato di interazioni complesse tra diversi attori. La regolamentazione della responsabilità giuridica per i danni causati dai sistemi di IA, pur rappresentando un punto di partenza significativo, presenta diverse criticità. Infatti, come già osservato, la Convenzione si limita a stabilire principi generali, la cui formulazione vaga rischia, anche in tal caso, di generare approcci disomogenei tra gli Stati, creando incertezza giuridica e frammentazione regolatoria. Il principio di responsabilità oggettiva, sebbene utile in astratto, non affronta adeguatamente la complessità delle interazioni tra i vari attori coinvolti nei sistemi di IA, lasciando irrisolta la questione della ripartizione che deriva dall'attribuzione di una responsabilità. Inoltre, l'assenza di meccanismi di armonizzazione tra le normative nazionali potrebbe ostacolare la certezza del diritto e la prevedibilità per gli operatori. Infine, la mancanza di un sistema di *enforcement* efficace e uniforme finisce per compromettere la concreta capacità di garantire la piena attuazione delle misure preventive e di responsabilità³⁷. In definitiva, il sistema di responsabilità e di *accountability* delineato appare incompleto, in assenza di strumenti chiari

³¹ *Ibidem*, art. 19: «Each Party shall seek to ensure that important questions raised in relation to artificial intelligence systems are, as appropriate, duly considered through public discussion and multistakeholder consultation in the light of social, economic, legal, ethical, environmental and other relevant implications».

³² *Ibidem*, art. 20: «Each Party shall encourage and promote adequate digital literacy and digital skills for all segments of the population, including specific expert skills for those responsible for the identification, assessment, prevention and mitigation of risks posed by artificial intelligence systems».

³³ *Ibidem*, art. 9: «Each Party shall adopt or maintain measures to ensure accountability and responsibility for adverse impacts on human rights, democracy and the rule of law resulting from activities within the lifecycle of artificial intelligence systems».

³⁴ *Ibidem*, Chapter III – Principles related to activities within the lifecycle of artificial intelligence systems: artt. 6-13.

³⁵ *Ibidem*, art. 16 par. 3: «Each Party shall adopt or maintain measures that seek to ensure that adverse impacts of artificial intelligence systems to human rights, democracy, and the rule of law are adequately addressed».

³⁶ *Ibidem*, art. 16, par. 1.

³⁷ *Ibidem*, art. 26: «1) Each Party shall establish or designate one or more effective mechanisms to oversee compliance with the obligations in this Convention; 2) Each Party shall ensure that such mechanisms exercise their duties independently and impartially and that they have the necessary powers, expertise and resources to effectively fulfil their tasks of overseeing compliance with the obligations in this Convention, as given effect by the Parties; 3) If a Party has provided for more than one such mechanism, it shall take measures, where practicable, to facilitate effective cooperation among them; 4) If a Party has provided for mechanisms different from existing human rights structures, it shall take measures, where practicable, to promote effective cooperation between the mechanisms referred to in paragraph 1 and those existing domestic human rights structures».

per garantire sanzioni proporzionate e uniformi da parte degli Stati contraenti, aggravato dalla mancanza di meccanismi di monitoraggio efficaci.

I meccanismi di monitoraggio individuati hanno l'obiettivo di valutare l'aderenza degli Stati agli *standard* della Convenzione³⁸ e includono l'istituzione di una Conferenza delle Parti³⁹, che avrà il compito di esaminare periodicamente la realizzazione delle disposizioni convenzionali e di formulare raccomandazioni per il loro miglioramento. Secondo alcuni, la presenza di tali meccanismi può risultare cruciale nel garantire non solo il rispetto formale delle norme, ma anche la loro effettiva applicazione nella pratica⁴⁰. Nella medesima prospettiva, la revisione periodica delle disposizioni della Convenzione consentirebbe di affrontare le sfide poste dalla rapida evoluzione tecnologica, assicurando che il quadro normativo possa rimanere aggiornato e pertinente⁴¹. Ad una lettura più attenta di queste disposizioni, si può, invece, rilevare che la mancanza di poteri vincolanti riduca l'efficacia delle raccomandazioni della Conferenza delle Parti, lasciando agli Stati un'eccessiva discrezionalità⁴². Inoltre, la Convenzione prevede la possibilità per quest'ultima di proporre al Comitato dei Ministri modalità adeguate al fine di coinvolgere esperti qualificati a supporto di un'attuazione efficace della Convenzione. Sebbene il monitoraggio indipendente sia pensato per garantire una supervisione imparziale, l'imposizione di valutazioni esterne potrebbe essere percepita dagli Stati come un'ingerenza, soprattutto in contesti in cui le normative nazionali divergono significativamente dagli standard della Convenzione⁴³. Infine, la revisione periodica delle disposizioni, pur necessaria per rispondere all'evoluzione tecnologica, rischia di risultare inefficace in assenza di procedure chiare per integrare rapidamente le modifiche nei sistemi giuridici nazionali⁴⁴. Senza un coordinamento strutturato tra gli Stati e la Conferenza delle Parti, il processo di aggiornamento regolatorio potrebbe rivelarsi lento o quantomeno parziale, compromettendo la capacità della Convenzione di rimanere al passo con i rapidi sviluppi dell'intelligenza artificiale.

Infine, la Convenzione si coordina con il quadro normativo in evoluzione dell'Unione europea, attribuendole un ruolo significativo: garantisce che l'applicazione della normativa dell'Unione relativa all'intelligenza artificiale avvenga in armonia con la Convenzione, evitando conflitti e riconosce esplicitamente la possibilità per l'Unione di aderire alla

³⁸ *Ibidem*, Chapter VII – Follow-up mechanism and co-operation: artt. 23-26.

³⁹ *Ibidem*, art. 23: «The Conference of the Parties shall be composed of representatives of the Parties to this Convention. The Parties shall consult periodically with a view to: a) facilitating the effective application and implementation of this Convention [...]; b) considering the possible supplementation to or amendment of this Convention; c) considering matters and making specific recommendations concerning the interpretation and application of this Convention; d) facilitating the exchange of information on significant legal, policy or technological developments of relevance [...]; e) facilitating, where necessary, the friendly settlement of disputes related to the application of this Convention; and f) facilitating co-operation with relevant stakeholders concerning pertinent aspects of the implementation of this Convention, including through public hearings where appropriate».

⁴⁰ A. FARANDA-G. LUSARDI, *Il primo trattato internazionale sull'IA: cosa prevede la convenzione quadro del Consiglio d'Europa*, in *Agenda Digitale*, 28 maggio 2024, reperibile online; F. NIOLA, *IA, ecco il primo trattato internazionale: la Convenzione del Consiglio d'Europa*, in *Agenda Digitale*, 6 settembre 2024, reperibile online; M. ODDO, *Intelligenza artificiale, verso regole globali: gli impatti della convenzione quadro*, in *Cybersecurity360*, 12 settembre 2024, reperibile online.

⁴¹ *Council of Europe Framework Convention on Artificial Intelligence and Human Rights*, cit., art. 23, par. 2, lett. a).

⁴² *Ibidem*, lett. c).

⁴³ *Ibidem*, par. 6: «The Conference of the Parties may propose to the Committee of Ministers appropriate ways to engage relevant expertise in support of the effective implementation of this Convention».

⁴⁴ *Ibidem*, art. 26.

Convenzione, diventandone formalmente Parte contraente⁴⁵. La Convenzione opera un richiamo al diritto dell'Unione⁴⁶, consentendo di segnalare violazioni che siano al contempo contrarie alla Convenzione e allo stesso diritto unionale, complessivamente inteso⁴⁷. L'espresso riferimento al diritto dell'Unione europea e alla giurisprudenza della Corte di giustizia, peraltro, riflette una volontà di coordinamento nello spazio giuridico europeo che rappresenta un *unicum*.

Anche in questo caso, tuttavia, ci sembra si pongano delle criticità non così facilmente superabili. Infatti, sebbene il riferimento esplicito al diritto dell'Unione e alla giurisprudenza della Corte di giustizia rappresenti un'innovazione, resta ambiguo il modo in cui saranno gestite eventuali discrepanze interpretative, creando potenziali incertezze normative. La possibilità che la Convenzione si configuri come un accordo misto, sottoscritto sia dall'Unione che dagli Stati membri, complica ulteriormente il quadro sul piano delle relazioni tra Unione e suoi Stati membri, in termini di competenze, sia sul piano dello stesso ordinamento UE che sul piano dei rapporti tra UE, Stati membri e Stati terzi europei contraenti la Convenzione, purché la stessa UE non voglia adottare una dichiarazione sulle competenze al momento dell'adesione formale. Inoltre, la Convenzione sembra mancare di un approccio realmente organico alla regolamentazione dell'IA, non affrontando in modo pienamente integrato le molteplici dimensioni coinvolte, come la protezione dei dati, la sicurezza e i diritti fondamentali. Infine, l'integrazione della Convenzione nell'*acquis* dell'Unione potrebbe generare incertezze sulla sua effettiva applicabilità nei procedimenti nazionali.

Dall'analisi che precede si evince come la Convenzione quadro del Consiglio d'Europa sull'intelligenza artificiale rappresenti sicuramente un passo avanti importante verso una regolamentazione internazionale delle tecnologie emergenti, ma sollevi interrogativi significativi sulla sua effettiva efficacia, nonché sulla sua capacità di bilanciare progresso tecnologico e tutela dei diritti fondamentali. Dall'analisi effettuata emergono forti dubbi sulla sua capacità concreta di mitigare i rischi dell'utilizzo dell'IA in assenza di obiettivi mirati, regolamentazione *ad hoc* o settoriale in favore di un approccio olistico che si dimostra utopistico. Inoltre, appare evidente come l'impostazione fortemente precauzionale del testo in esame rischi di limitare lo sviluppo dell'IA, senza offrire garanzie sufficienti per affrontare le sfide globali poste da queste tecnologie.

⁴⁵ Un passo importante in tal senso è stato compiuto con la firma della Convenzione da parte della Commissione europea il 5 settembre 2024, un gesto che sottolinea non solo la volontà dell'Unione ma anche dei suoi Stati membri di vincolarsi a tale strumento. L'adesione formale dell'Unione renderebbe la Convenzione parte dell'*acquis* comunitario, consentendone l'applicazione anche nei procedimenti giudiziari degli Stati membri. Paradossalmente, l'intera procedura potrebbe concludersi più rapidamente rispetto al percorso, ancora in sospenso, per l'adesione dell'Unione alla Convenzione europea dei diritti dell'uomo. Si veda in proposito G. ZACCARONI, *Intelligenza artificiale e principio democratico*, cit., 11-22.

⁴⁶ *Council of Europe Framework Convention on Artificial Intelligence and Human Rights*, cit., art. 27, par. 2: «Parties which are members of the European Union shall, in their mutual relations, apply European Union rules governing the matters within the scope of this Convention without prejudice to the object and purpose of this Convention and without prejudice to its full application with other Parties. The same applies to other Parties to the extent that they are bound by such rules».

⁴⁷ Cfr. *Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, 5 September 2024, CE/TS 225, par. 147: «[...] measures, principles and procedures provided for in the European Union legal order, in particular laws, regulations or administrative provisions as well as other requirements, including court decisions [...]».

4. Il modello normativo dell'Unione europea sull'IA: tra idealismo regolatorio e vincoli applicativi

La regolamentazione dell'Unione Europea in materia di IA si colloca all'interno di una strategia ampia e ambiziosa volta a bilanciare l'innovazione tecnologica con la tutela dei diritti fondamentali, della sicurezza e della trasparenza. Come è noto, al centro di questa iniziativa normativa vi è il Regolamento sull'Intelligenza Artificiale (*AI Act*)⁴⁸, proposto dalla Commissione Europea nell'aprile 2021, che rappresenta il primo tentativo organico di disciplinare l'IA⁴⁹. Oltre all'*AI Act*, la materia è integrata da altre iniziative legislative europee, come il Regolamento generale sulla protezione dei dati (GDPR)⁵⁰, che stabilisce standard elevati per il trattamento dei dati personali utilizzati dalle tecnologie di IA, il *Digital Services Act* (DSA)⁵¹ e il *Digital Markets Act* (DMA)⁵², che affrontano le responsabilità delle piattaforme digitali e delle *big tech*, promuovendo un ambiente digitale sicuro e competitivo⁵³. Da ultimo, si deve segnalare che è in corso di preparazione una direttiva relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'IA⁵⁴. La Commissione Europea ha anche lanciato il Piano Coordinato sull'IA⁵⁵, che mira a promuovere la cooperazione tra Stati membri per lo sviluppo e l'implementazione dell'IA, e il programma *Horizon Europe*⁵⁶, che finanzia la ricerca e l'innovazione nel campo dell'intelligenza artificiale.

L'analisi che segue non ha come obiettivo di focalizzarsi sulle specifiche soluzioni giuridiche elaborate dagli strumenti richiamati, oggetto di approfondite riflessioni in altri contributi di questo *Focus*, ma di evidenziarne gli aspetti di criticità che emergono, sempre

⁴⁸ Regulation (EU) 2024/1689 Of The European Parliament and of the Council of 13 June 2024, cit.

⁴⁹ Il regolamento su ll'intelligenza artificiale (*AI Act*) è stato estesamente commentato. Si vedano da ultimo i contributi inseriti nei *Quaderni AISDUE* - Rivista quadrimestrale ISSN 2975-2698 - Fascicolo speciale n. 2/2024 La nuova disciplina UE sull'intelligenza artificiale, e in particolare: E. CIRONE, *L'AI Act e l'obiettivo (mancato?) di promuovere uno standard globale per la tutela dei diritti fondamentali*; F. FERRI, *Il giorno dopo la rivoluzione: prospettive di attuazione del regolamento sull'intelligenza artificiale e poteri della Commissione Europea*; M. INGLESE, *Il regolamento sull'intelligenza artificiale come atto per il completamento e il buon funzionamento del mercato interno?*; S. VILLANI, *Il sistema di vigilanza sull'applicazione dell'AI Act: ognun per sé?*; nonché C. NOVELLI, *L'Artificial Intelligence Act europeo: alcune questioni di implementazione*, in *federalismi.it*, 2024, 95 ss.; C. NOVELLI-F. CASOLARI-A. ROTOLO-M. TADDEO-L. FLORIDI, *Taking AI risks seriously: a new assessment model for the AI Act*, in *AI & Society*, 2023, 1 ss.; J. SCHUETT, *Risk Management in the Artificial Intelligence Act*, in *European Journal of Risk Regulation*, 2023, 1 ss.; M. ALMADA-N. PETT, *The EU AI Act: a Medley of Product Safety and Fundamental Rights?*, in *Robert Schuman Centre Working Paper*, 2023, reperibile online.

⁵⁰ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119, 4.5.2016, Corrigendum, OJ L 127, 23.5.2018.

⁵¹ Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), L 277/1, 27.10.2022.

⁵² Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act), L 265/1, 12.10.2022.

⁵³ Per un'analisi sistematica degli strumenti in esame si veda C. SCHEPISI, *Le "dimensioni" della regolazione dell'intelligenza artificiale nella proposta di regolamento della Commissione*, in *Quaderni AISDUE*, Sezione "Atti convegni AISDUE", 16, 2022, 347 ss.

⁵⁴ Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (*AI Liability Directive*), COM/2022/496 final, reperibile online.

⁵⁵ Commissione Europea, *Piano coordinato sull'intelligenza artificiale*, COM/2018/795 final, reperibile online.

⁵⁶ European Commission: Directorate-General for Research and Innovation, *Horizon Europe, open science – Early knowledge and data sharing, and open collaboration*, Publications Office of the European Union, 2021, reperibile online.

con il fine di mostrare come anche il modello unionale, per quanto pionieristico, risenta di impostazioni logico-giuridiche a tratti utopistiche e si scontri nella sua applicazione con ostacoli sistemici difficilmente superabili⁵⁷.

In tale prospettiva, dopo aver effettuato alcune considerazioni preliminari in merito alla finalità, all'approccio e alla definizione stessa di intelligenza artificiale contenuta nell'*AI Act*, ci si concentrerà sui due aspetti chiave dello strumento in esame: la questione della tutela dei diritti umani e l'approccio basato sul rischio.

L'esame complessivo del Regolamento mostra innanzitutto l'intenzione del legislatore dell'Unione europea di creare una regolamentazione con un raggio d'azione che si estenda oltre i confini unionali e, in secondo luogo, la volontà di adottare volutamente un approccio orizzontale, che non si concentri su problematiche settoriali o specifiche, ma miri a disciplinare trasversalmente tutti i settori interessati dall'intelligenza artificiale. A ciò si aggiunga l'incerta formulazione della definizione di IA volta a ricomprendere un ampio spettro di usi della stessa.

Sia l'obiettivo perseguito che l'approccio adottato risultano discutibili. Quanto all'ambizione di proiettare il quadro normativo oltre i confini dell'Unione, pur rappresentando un tentativo di acquisire una posizione di "supremazia normativa" a livello globale, rischia di rivelarsi problematica. La regolamentazione UE, ispirata da valori come la protezione dei diritti fondamentali e la sicurezza, potrebbe entrare in conflitto con le priorità normative e commerciali di altre giurisdizioni, come quelle di Stati Uniti, Cina e India, che adottano approcci più flessibili o orientati alla promozione dell'innovazione, creando tensioni economiche e politiche e ostacolando la cooperazione globale in questo ambito. Inoltre, l'imposizione di standard UE a operatori esterni potrebbe indurre alcune imprese a evitare il mercato europeo, con conseguenti ripercussioni negative sulla competitività dell'industria tecnologica nell'UE.

Con riferimento all'adozione di un approccio orizzontale, volto a disciplinare in modo uniforme tutti i settori interessati dall'intelligenza artificiale, quest'ultimo pur garantendo coerenza normativa, si scontra con la complessità e la diversità delle applicazioni dell'IA. Questo approccio, ispirato al modello del GDPR, mira a evitare frammentazioni normative tra gli Stati membri e a garantire che ogni utilizzo dell'IA rispetti standard comuni di tutela dei diritti fondamentali. Tuttavia, questa impostazione solleva interrogativi circa la sua efficacia pratica. La pluralità degli ambiti di applicazione dell'IA rende difficile stabilire requisiti univoci che siano al contempo adeguati e sufficientemente specifici⁵⁸. Inoltre, la regolamentazione orizzontale rischia di creare sovrapposizioni con altre normative settoriali, generando ulteriore complessità non solo per gli operatori giuridici, ma soprattutto per gli operatori del mercato. Infine, l'integrazione di normative IA con i regolamenti già esistenti (ad esempio nel settore sanitario o dei trasporti) richiederà un coordinamento significativo tra le autorità nazionali e sovranazionali, con il rischio di rallentare l'attuazione delle nuove regole.

Quanto al concetto stesso di intelligenza artificiale, quest'ultima è concepita in chiave metagiuridica come «una famiglia di tecnologie in rapida evoluzione che contribuisce al conseguimento di un'ampia gamma di benefici a livello economico, ambientale e sociale

⁵⁷ C. NOVELLI, *L'Artificial Intelligence Act Europeo*, cit., 102-105.

⁵⁸ Ad esempio, l'art. 13 dell'*AI Act* stabilisce obblighi generali di trasparenza per tutti i sistemi di IA, ma non tiene conto delle differenze operative tra una piattaforma di *e-commerce* e un sistema di sorveglianza pubblico: «High-risk AI systems shall be designed and developed in such a way as to ensure that their operation is sufficiently transparent to enable deployers to interpret a system's output and use it appropriately [...]».

nell'intero spettro delle attività industriali e sociali»⁵⁹. Si tratta di una definizione, ampia e generica, che riflette la natura fluida del concetto stesso di IA e l'intento di adattarsi ai rapidi sviluppi tecnologici e alla varietà delle sue applicazioni. Tuttavia, tale scelta definitoria sebbene abbia il pregio di riconoscere la natura eterogenea e dinamica dell'IA, solleva inevitabilmente difficoltà nella traduzione pratica dei principi normativi. In secondo luogo, questa impostazione metagiuridica sembra confermare l'intenzione dell'Unione Europea di creare un quadro normativo universale, capace di adattarsi all'evoluzione costante della tecnologia. Sebbene questa flessibilità possa sembrare vantaggiosa per affrontare il carattere dinamico dell'IA, essa comporta il rischio di un'applicazione normativa frammentata o incoerente nella prassi degli Stati membri. Inoltre, la scelta di una definizione così inclusiva, pur ponendosi l'obiettivo di coprire un ampio spettro di applicazioni, tende a diluire l'attenzione su questioni specifiche e critiche, come la trasparenza degli algoritmi, la protezione dei dati personali, la responsabilità giuridica per danni causati dall'IA e la prevenzione di discriminazioni. Infine, l'adozione di una definizione così generale riflette anche la complessità di elaborare una normativa che possa essere applicabile non solo all'interno dell'Unione, ma anche a livello globale.

Quanto ai principali aspetti del Regolamento, la preoccupazione per la possibilità di compressione dei diritti fondamentali e l'approccio basato sul rischio ne costituiscono i pilastri fondamentali.

Innanzitutto, come la Convenzione del Consiglio d'Europa, anche lo strumento in esame pone l'accento sulla protezione dei diritti fondamentali, considerandola l'obiettivo principale della normativa⁶⁰. Questo principio emerge chiaramente nell'articolato del Regolamento, dove l'importanza della salvaguardia della dignità umana, della *privacy* e della non discriminazione viene ribadita più volte, in linea con gli *standard* già stabiliti in altri strumenti del diritto dell'Unione Europea⁶¹, come la Carta dei Diritti Fondamentali dell'UE⁶² e lo stesso Regolamento Generale sulla Protezione dei Dati (GDPR)⁶³. Ampio spazio peraltro è riservato al problema della trasparenza. Quest'ultima è intesa come la capacità di garantire che i sistemi di IA siano progettati e utilizzati in modo tracciabile e spiegabile, informando adeguatamente sia gli utenti finali che i responsabili del loro impiego, inclusa la

⁵⁹ *Ibidem*, Considerando 4: «AI is a fast evolving family of technologies that contributes to a wide array of economic, environmental and societal benefits across the entire spectrum of industries and social activities [...]».

⁶⁰ A. ADINOLFI, *L'unione europea dinanzi allo sviluppo dell'intelligenza artificiale: la costruzione di uno schema di regolamentazione europeo tra mercato unico digitale e tutela dei diritti fondamentali*, in S. DORIGO (a cura di), *Il ragionamento giuridico nell'era della intelligenza artificiale*, Pisa, 2020, 14 ss.

⁶¹ Si vedano sul punto le riflessioni di G. CONTALDI, *La proposta di regolamento sull'intelligenza artificiale e la protezione dei dati personali*, in G. CAGGIANO-G. CONTALDI-P. MANZINI (a cura di), *Verso una legislazione europea su mercati e servizi digitali*, Bari, 2021, 207 ss.; ma anche C. NOVELLI-F. CASOLARI-P. HACKER-G. SPEDICATO-L. FLORIDI, *Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity*, in *Computer Law and Security Review*, 55, 2024, 1-16.

⁶² Gazzetta ufficiale dell'Unione europea, *Carta dei diritti fondamentali dell'Unione europea*, 2016/C 202/02, 389-405. L'utilizzo dell'IA con le sue caratteristiche specifiche (ad esempio opacità, complessità, dipendenza dai dati, comportamento autonomo) può incidere negativamente su una serie di diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea. Si veda in proposito Commissione Europea, *Proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione – Relazione*, COM/2021/206 final, punto 3.5; Commissione Europea, *Creare fiducia nell'intelligenza artificiale antropocentrica*, COM/2019/168 final; *Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate*, 2020/2012(INL).

⁶³ *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016*, cit., art. 1.

comunicazione sui limiti e le potenzialità del sistema, oltre ai diritti delle persone coinvolte. Il Regolamento, infine, amplia il discorso oltre la protezione dei diritti fondamentali, assumendo una prospettiva globale che collega l'intelligenza artificiale a pratiche commerciali scorrette, come definite nella direttiva 2005/29/CE⁶⁴. Nel Considerando 29⁶⁵, si evidenzia come l'IA possa essere utilizzata per manipolare le persone, inducendole ad adottare comportamenti indesiderati o a prendere decisioni in modo ingannevole, compromettendo la loro autonomia e capacità decisionale.

Uno degli elementi distintivi dell'*AI Act* è l'adozione di misure specificamente progettate per mitigare i rischi che l'intelligenza artificiale può rappresentare per i diritti fondamentali⁶⁶. In questo senso, un aspetto fondamentale della normativa è l'adozione del cosiddetto approccio basato sul rischio⁶⁷, simile a quello utilizzato in altre normative europee, come il GDPR⁶⁸ o gli atti normativi in materia di antiriciclaggio⁶⁹. Tale approccio è utilizzato per classificare i sistemi di IA in base al loro potenziale di arrecare danno ai diritti fondamentali, sottoponendo ciascuno di essi a regimi normativi differenti. L'art. 6 distingue i sistemi ad alto rischio, che includono applicazioni in settori come la sanità, la giustizia e il lavoro, e quelli a rischio inaccettabile⁷⁰, come i sistemi di manipolazione subliminale o di sorveglianza biometrica in tempo reale, vietati dall'art. 5⁷¹. Alcuni sistemi di IA sono vietati all'interno dell'Unione perché considerati lesivi della dignità umana o di diritti fondamentali, mentre altri sono classificati come ad alto rischio sulla base di criteri specifici. Tra questi, si annoverano: lo scopo del sistema di IA, la quantità e la natura dei dati trattati, il grado di autonomia del sistema, la possibilità di intervento umano per mitigare decisioni potenzialmente dannose, la probabilità e l'entità dei danni causati e la vulnerabilità delle persone coinvolte, in relazione a fattori come età, condizione sociale o economica, o squilibri di potere rispetto agli utilizzatori del sistema.

⁶⁴ Direttiva 2005/29/CE del Parlamento europeo e del Consiglio dell'11 maggio 2005 relativa alle pratiche commerciali sleali tra imprese e consumatori nel mercato interno e che modifica la direttiva 84/450/CEE del Consiglio e le direttive 97/7/CE, 98/27/CE e 2002/65/CE del Parlamento europeo e del Consiglio e il regolamento (CE) n. 2006/2004 del Parlamento europeo e del Consiglio («direttiva sulle pratiche commerciali sleali»), L 149/22, 11.6.2005.

⁶⁵ Regulation (EU) 2024/1689 Of The European Parliament and of the Council of 13 June 2024, cit., Considerando 29: «AI-enabled manipulative techniques can be used to persuade persons to engage in unwanted behaviours, or to deceive them by nudging them into decisions in a way that subverts and impairs their autonomy, decision-making and free choices».

⁶⁶ C. SCHEPISI, *Diritti fondamentali, principi democratici e rule of law: quale ruolo e quale responsabilità per gli Stati nella regolazione dell'intelligenza artificiale*, in *Studi sull'integrazione europea*, 2022, 41-66.

⁶⁷ Su tale sistema, e sulle differenze e gradazioni nei diversi atti legislativi dell'Unione che lo applicano, vedi G. DE GREGORIO-P. DUNN, *The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age*, in *Common Market Law Review*, 2022, 473-500.

⁶⁸ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, cit., art. 35: Data protection impact assessment.

⁶⁹ Regolamento (UE) 2024/1620 del Parlamento europeo e del Consiglio del 31 maggio 2024 che istituisce l'Autorità per la lotta al riciclaggio e al finanziamento del terrorismo e che modifica i regolamenti (UE) n. 1093/2010, (UE) n. 1094/2010 e (UE) n. 1095/2010, GU L del 19.6.2024, art. 8: Metodologia di supervisione AML/CFT.

⁷⁰ Regulation (EU) 2024/1689 Of The European Parliament and of the Council of 13 June 2024, cit., art. 6: «Irrespective of whether an AI system is placed on the market or put into service independently of the products referred to in points (a) and (b), that AI system shall be considered to be high-risk where both of the following conditions are fulfilled: (a) the AI system is intended to be used as a safety component of a product, or the AI system is itself a product, covered by the Union harmonization legislation listed in Annex I; (b) the product whose safety component pursuant to point (a) is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment, with a view to the placing on the market or the putting into service of that product pursuant to the Union harmonization legislation listed in Annex I».

⁷¹ *Ibidem*, art. 5, par. 1, lett. a), b).

Sebbene questa impostazione consenta una regolamentazione flessibile, essa pone alcuni dubbi, soprattutto di ordine applicativo. Innanzitutto, la definizione stessa di “rischio” è problematica. L’ampiezza della definizione dei sistemi di IA, che include tutte le tecnologie basate su apprendimento automatico, ragionamento logico e statistico, rischia di generare incertezze interpretative. Questo potrebbe portare a discrepanze nell’applicazione del regolamento tra gli Stati membri, vanificando l’obiettivo di un’armonizzazione normativa. Inoltre, il Regolamento non chiarisce sufficientemente i criteri per determinare il livello di rischio. L’art. 14, ad esempio, impone che i sistemi ad alto rischio siano soggetti a supervisione umana per mitigare i rischi imprevisti, ma non specifica quali siano gli standard per valutare tali rischi⁷². Questa lacuna evidenzia un paradosso intrinseco: mentre l’approccio precauzionale mira a prevenire i danni, l’incertezza normativa potrebbe frenare l’innovazione, scoraggiando le imprese dall’investire in tecnologie IA all’interno dell’Unione Europea. Infine, il Regolamento prevede che i sistemi di IA che non rientrano nelle categorie ad alto rischio non siano soggetti a obblighi stringenti, a meno che non presentino rischi significativi per la salute, la sicurezza o i diritti fondamentali, o non influenzino in maniera rilevante l’esito di processi decisionali. Questo approccio graduato mira a bilanciare l’innovazione tecnologica con la protezione degli individui, evitando un’applicazione normativa eccessivamente onerosa per i sistemi a rischio limitato.

Il Regolamento introduce, inoltre, la valutazione di impatto sui diritti fondamentali come elemento chiave per garantire che l’introduzione e l’utilizzo dei sistemi di IA siano conformi agli *standard* etici e legali europei. Ai sensi del primo paragrafo dell’art. 27, la valutazione di impatto è articolata in sei sezioni principali: a) una descrizione dei processi nei quali il sistema verrà impiegato, b) una definizione del tempo e della frequenza di utilizzo, c) l’individuazione delle categorie di persone fisiche e gruppi che saranno potenzialmente interessati nel contesto specifico, d) l’analisi dei rischi specifici per tali categorie, e) una spiegazione delle modalità di sorveglianza umana previste, e f) le misure progettate per mitigare i rischi identificati⁷³.

Tuttavia, anche questo sistema presenta alcune criticità che meritano attenzione. Innanzitutto, non è chiaro il rapporto tra questa valutazione di impatto e il più ampio sistema di gestione del rischio previsto dal Regolamento. Sebbene il secondo paragrafo dell’art. 27 sembri indicare che si tratti di due processi distinti⁷⁴, esiste il rischio di sovrapposizioni o di confusione tra i vari obblighi, specialmente quando il *deployer* deve integrare valutazioni già effettuate dal fornitore o previste da altre normative. Questa complessità potrebbe aumentare gli oneri amministrativi per gli operatori, soprattutto per quelli meno strutturati, come le

⁷² *Ibidem*, art. 14: «High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which they are in use. Human oversight shall aim to prevent or minimize the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse [...]».

⁷³ *Ibidem*, art. 27, par. 1: «Prior to deploying a high-risk AI system referred to in Article 6(2), with the exception of high-risk AI systems intended to be used in the area listed in point 2 of Annex III, deployers that are bodies governed by public law, or are private entities providing public services, and deployers of high-risk AI systems referred to in points 5 (b) and (c) of Annex III, shall perform an assessment of the impact on fundamental rights that the use of such system may produce [...]».

⁷⁴ *Ibidem*, par. 2: «The obligation laid down in paragraph 1 applies to the first use of the high-risk AI system. The deployer may, in similar cases, rely on previously conducted fundamental rights impact assessments or existing impact assessments carried out by provider. If, during the use of the high-risk AI system, the deployer considers that any of the elements listed in paragraph 1 has changed or is no longer up to date, the deployer shall take the necessary steps to update the information».

piccole e medie imprese. Un'ulteriore criticità riguarda l'effettiva qualità e uniformità delle valutazioni di impatto. L'assenza di linee guida dettagliate su come condurre queste valutazioni potrebbe portare a interpretazioni divergenti tra i diversi operatori o Stati membri, compromettendo l'obiettivo di coerenza e armonizzazione del Regolamento. Inoltre, il rischio che le valutazioni siano trattate come meri adempimenti burocratici, piuttosto che come strumenti per identificare e mitigare rischi reali, potrebbe limitarne l'efficacia pratica. Un altro elemento problematico è legato alla notifica dei risultati all'autorità di vigilanza del mercato. Sebbene questa misura miri a rafforzare la trasparenza e la supervisione, non è chiaro se le autorità disporranno delle risorse e delle competenze necessarie per analizzare e verificare un numero potenzialmente elevato di valutazioni. Ciò potrebbe creare ritardi o disfunzioni operative, riducendo l'impatto concreto della regolamentazione. Infine, il quinto paragrafo dell'art. 27 assegna all'Ufficio per l'Intelligenza Artificiale il compito di predisporre un questionario semplificato per agevolare i *deployer*⁷⁵, sebbene resti incerta l'effettiva utilità di questo strumento, poiché la semplificazione potrebbe compromettere la capacità di affrontare rischi complessi.

Il Regolamento delinea in definitiva un quadro giuridico complesso, centrato sulla protezione dei diritti fondamentali attraverso la gestione del rischio e la valutazione dell'impatto. Ci sembra, tuttavia, che sia necessario operare qualche riflessione, sia sul piano concettuale che pratico.

In primo luogo, come già rilevato l'approccio adottato risulta eccessivamente generalista. La scelta di disciplinare l'intero ecosistema dell'intelligenza artificiale con un unico quadro regolatorio, senza considerare le peculiarità dei vari settori, rischia di trasformare il Regolamento in una raccolta di enunciazioni di principio, più che in un insieme di norme chiare, precise ed immediatamente applicabili senza ulteriori precisazioni. La varietà delle applicazioni dell'IA richiederebbe una regolamentazione più specifica e settoriale, in grado di rispondere alle sfide proprie di ciascun contesto tecnologico e sociale. La mancanza di tale differenziazione non solo complica l'applicazione pratica delle norme, ma aumenta il rischio di interpretazioni divergenti tra gli Stati membri, compromettendo l'armonizzazione normativa all'interno dell'Unione.

Infine, l'approccio precauzionale del Regolamento, che impone rigidi obblighi di conformità e complessi meccanismi di valutazione del rischio, rischia di soffocare lo sviluppo tecnologico delle aziende europee. Sebbene l'intento dell'Unione sia quello di estendere l'efficacia normativa del Regolamento oltre i confini europei, con una portata extraterritoriale, nella pratica questa ambizione potrebbe rivelarsi controproducente. Le imprese europee, costrette a operare in un contesto normativo stringente e burocratizzato, si troverebbero in una posizione di svantaggio competitivo rispetto a quelle che operano in giurisdizioni con regole più flessibili, come Stati Uniti, Cina o India. Questo scenario rischia di penalizzare l'innovazione e gli investimenti nel settore tecnologico europeo, incentivando potenzialmente una delocalizzazione delle attività verso contesti regolatori meno restrittivi.

In definitiva, il Regolamento, pur avendo l'obiettivo di garantire la tutela dei diritti fondamentali e la sicurezza nell'uso dell'intelligenza artificiale, rischia di fallire nel bilanciare adeguatamente queste priorità con la necessità di promuovere lo sviluppo tecnologico e la competitività dell'industria europea. Anche in UE, ci sembra che un approccio più mirato,

⁷⁵ *Ibidem*, par. 5: «The AI Office shall develop a template for a questionnaire, including through an automated tool, to facilitate deployers in complying with their obligations under this Article in a simplified manner».

flessibile e calibrato sui diversi contesti applicativi potrebbe offrire una soluzione più efficace per affrontare le sfide dell'IA senza compromettere il progresso tecnologico.

5. Conclusioni

La regolamentazione dell'intelligenza artificiale si colloca oggi al crocevia tra diritto, tecnologia e valori fondamentali, ponendo il giurista di fronte a una sfida complessa e in continua evoluzione. Le riflessioni sviluppate nel presente contributo hanno evidenziato come gli attuali modelli normativi, seppur significativi nel tentativo di affermare *standard* etici condivisi, risultino ancora caratterizzati da una tensione irrisolta tra esigenze di tutela e promozione dell'innovazione.

Tanto la Convenzione quadro del Consiglio d'Europa quanto il Regolamento (UE) 2024/1689 testimoniano l'intento di offrire una disciplina organica del fenomeno, fondata su un approccio precauzionale e antropocentrico. Tuttavia, tali strumenti rivelano una tendenza all'universalizzazione e alla rigidità normativa che rischia di comprometterne l'effettività applicativa, soprattutto in contesti eterogenei e ad alta variabilità tecnologica. L'adozione di un'impostazione olistica, se da un lato consente di affermare principi comuni di tutela, dall'altro mostra evidenti limiti operativi: le definizioni ampie di intelligenza artificiale, l'assenza di meccanismi adattivi e la scarsa articolazione per ambiti d'uso determinano incertezza interpretativa e applicativa. Inoltre, la concentrazione sulla mitigazione del rischio *ex ante*, se non accompagnata da adeguati strumenti di governance settoriale e di cooperazione internazionale, rischia di tradursi in un effetto paralizzante sul piano dell'innovazione, in particolare per gli attori meno strutturati.

In questa prospettiva, risulta necessario superare la logica ipertrofica e orizzontale che caratterizza le attuali proposte normative, orientandosi verso modelli funzionali, settoriali e proporzionati, capaci di garantire un bilanciamento più efficace tra i diversi interessi in gioco. Una regolazione giuridica realmente adeguata al fenomeno dell'IA dovrebbe essere non solo fondata su principi, ma anche dotata di strumenti adattivi, differenziati e capaci di cogliere la specificità dei contesti applicativi.

In definitiva, la regolazione dell'intelligenza artificiale richiede un ripensamento dell'attuale impianto normativo, che, nella sua forma più recente, continua a privilegiare modelli orizzontali e generalisti. Sebbene tali approcci rispondano all'esigenza di affermare principi etici condivisi a livello internazionale, essi si rivelano, nella pratica, poco adatti a governare efficacemente la varietà e la complessità delle applicazioni tecnologiche. Una disciplina giuridica realmente efficace dovrebbe fondarsi su strumenti normativi flessibili, graduati e capaci di adattarsi alle peculiarità dei singoli contesti di impiego, senza sacrificare la coerenza del quadro complessivo. In questo senso, la sfida non è tanto quella di produrre nuove norme, quanto di affinare gli strumenti esistenti attraverso soluzioni tecniche, procedurali e istituzionali che permettano un equilibrio più solido tra tutela e innovazione.



ROBERTO AMAGLIANI*

INTELLIGENZA ARTIFICIALE ED INTERESSE SUPERIORE DEL MINORE

Sommario: 1. Premessa – 2. Il regolamento UE sulla I.A. e la tutela del minore – 3. La deriva procedimentale/processualistica: la matrice prevalentemente giudiziale della determinazione dell'interesse del minore – 4. La insoddisfacente prospettiva della contrapposizione rischi/opportunità per adeguatamente contestualizzare la posizione (e le prospettive di realizzazione dell'interesse) del minore nell'era dell'intelligenza artificiale – 5. Il caso Replika.

1. Premessa

Il tema trattato propone l'accostamento di due termini, Intelligenza Artificiale¹ e Interesse del minore, la cui definizione è altamente problematica², ma, nel contempo,

* Professore ordinario di Diritto privato, Università di Messina.

¹ Come noto, l'evoluzione e lo sviluppo dei temi e delle dinamiche legati all'I. A. si deve essenzialmente alla cd. "informatica non tradizionale", nel cui ambito si muove la ricerca volta allo studio ed alla creazione di sistemi progettati in modo da avere quelle stesse caratteristiche che si associano all'intelligenza umana intesa come complesso di facoltà psichiche e mentali che consentono all'uomo di pensare, comprendere o spiegare i fatti o le azioni, elaborare modelli astratti della realtà, intendere e farsi intendere dagli altri, giudicare, e lo rendono insieme capace di adattarsi a situazioni nuove e di modificare la situazione stessa quando questa presenta ostacoli all'adattamento. E pertanto nell'ambito di tale ricerca gli obiettivi perseguiti sono essenzialmente due: - approfondire e comprendere i principi che rendono possibile l'intelligenza; - progettare computer dotati di capacità simili a quelle umane senza, però, tentare di imitare esattamente i processi informativi degli esseri umani, e cioè le modalità di apprendimento da parte dell'uomo, che nel caso dell'I.A. sono sostituite dalla predisposizione e dall'utilizzazione di algoritmi.

² Quella del testo è constatazione assolutamente scontata e riguarda in pari misura entrambi i profili presi in considerazione. Gli studiosi dell'intelligenza artificiale mettono infatti in luce come non vi sia *in radice* una definizione condivisa di "intelligenza" (per un'ampia trattazione di questo profilo v. G. SARTOR, *L'intelligenza artificiale ed il diritto*, Torino, 2022, 1 ss.) e nel contempo segnalano, sin dal secolo scorso, come "dell'intelligenza artificiale possono essere date varie definizioni, che però finiscono inevitabilmente per risultare o troppo generiche o troppo restrittive" (così, DELL'ORCO e POGGIO, *Intelligenza artificiale*, in *Enciclopedia del novecento, Supplemento*, vol. VIII, Milano, 1989) e, nello stesso senso, ora Sartor, *op. cit.*, 7 ss. Per quel che concerne la (controversa) definizione della categoria dell'interesse superiore del minore sia permesso di fare rinvio, per una rassegna delle tesi avanzate al riguardo, a R. AMAGLIANI, *La nuova disciplina della filiazione (etichette, formule magiche e principi nel diritto di famiglia)*, in *Giust. Civ.*, 2018, 1023 ss.; per una prospettiva assai critica nei confronti di quella

l'argomento prescelto invita alla inevitabile riflessione sulla possibile, e per molte ragioni auspicabile, incidenza/interferenza del diritto sugli sviluppi della tecnologia nuova³, seppur avuto riguardo ad uno specifico ambito dell'ordinamento civile.

Sul primo versante il regolamento dell'U. E. sull'I. A.⁴, al fine di tracciare l'ambito di applicabilità della disciplina ivi contenuta, propone, all'art. 3, un enunciato caratterizzato da un alto grado di generalità/genericità, ma al tempo stesso ispirato da un chiaro intento di onnicomprensività, riconducendo l'I.A. a qualsivoglia «sistema automatizzato progettato per funzionare con livelli di autonomia variabile 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, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali».

Per i nostri fini è soprattutto l'ultima proposizione che, come vedremo più avanti, mostra particolare importanza e deve essere quindi sin d'ora opportunamente sottolineata.

Venendo all'altro polo della dicotomia, se in virtù di enunciati normativi apicali univoci e diffusi, l'interesse del minore gode di sicuro rilievo ordinamentale e di grado assiologico prevalente⁵, non si può invece con altrettanta certezza affermarsene la qualificazione in termini di principio generale, piuttosto che di categoria dogmatica, o di clausola generale

che viene individuata quale *communis opinio* di matrice “minoricentrica”, cfr. di recente A. SPADAFORA, *La genitorialità responsabile, l'empito copernicano e l'eliocentrismo minorile*, in *Dir. fam. pers.*, 2025, 321 ss.

³ Tra i numerosi contributi in argomento si può consultare, oltre a G. SARTOR, *op. cit.*, *passim*, anche G. DI ROSA, *Categorie giuridiche e innovazione tecnologica*, in *Diritto e tecnologia. Precedenti storici e problematiche attuali*, Atti delle giornate di studi (Catania, 8 ottobre 2021 - 21 e 22 ottobre 2022 25 novembre 2022 - 19 e 20 maggio 2023), Napoli, 2024, 285 ss.; M. PALAZZO, *Potere della tecnica e funzione del diritto*, in *Biblioteca della fondazione italiana del notariato*, fasc.1, pag. 115 ss.; AA. VV., *Cyberspazio e diritto*, in *Rivista internazionale di informatica giuridica*, vol 24 n. 75 (3-2023), *passim*; AA. VV., *Diritto e intelligenza artificiale*, a cura di G. CASONATO, M. FASAN E S. PENASA, in *DPCE online*, sezione monografica, fasc. 1/2022. Si deve osservare, in generale, che nel settore in esame, una ormai risalente stratificazione normativa si affianca a svolgimenti ordinamentali che si succedono repentinamente e che nel contempo sono caratterizzati da una evidente *provisorietà* (come dimostra proprio il Reg. U. E. 1689/2024 il cui reale valore prescrittivo deve essere misurato comparando i quasi duecento “considerando” che ne costituiscono la premessa con la metà circa degli articoli di cui il regolamento è composto) così da escludere l'esattività e la proficuità di un studio limitato all'esistente (o allo stato dell'arte) in un momento storico dato, al quale, per lungo tempo, la dottrina è stata quasi esclusivamente adusa dedicarsi. Questa riflessione dà conto poi, dal punto di vista della politica del diritto (anche questa è prospettiva largamente trascurata dagli studi tradizionali più risalenti), di due esigenze che, nel caso dell'I.A., per qualche verso si contrappongono e si autolimitano. Da un lato, infatti, sussiste il timore che una legislazione prematura ed invadente possa ostacolare il progresso scientifico o, peggio ancora, causare inefficienze economiche. Allo stesso tempo si ammette che la mancanza di un ambiente giuridico affidabile e sicuro possa ugualmente ostacolare l'innovazione tecnologica.

⁴ Per un primo commento al regolamento U. E. si può consultare, AA. VV., *AI Act. Principi, regole ed applicazioni pratiche del Reg. Ue 1689/2024*, a cura di M. IASELLI, Rimini, 2024.

⁵ Come noto mentre la Costituzione italiana non contiene riferimenti espliciti all'interesse superiore del minore, ciò avviene invece nella Convenzione sui diritti del fanciullo fatta a New York il 20.11.1989 e ratificata dall'Italia con la L. 27 maggio 1991 n. 176, all'art. 3 1° comma, nella Convenzione europea sull'esercizio dei diritti dei fanciulli, fatta a Strasburgo il 25 gennaio 1996 e ratificata nel nostro Paese con L. 20 marzo 2003 n. 77, all'art. 1 1° comma, ed infine nella Carta dei diritti fondamentali dell'Unione Europea, all'art. 24 cpv, a mente del quale «In tutti gli atti relativi ai minori, siano essi compiuti da autorità pubbliche o da istituzioni private l'interesse del minore deve essere considerato preminente». Non si deve tuttavia trascurare che il nostro legislatore sembra maggiormente incline ad usare sinonimi dell'espressione “interesse superiore del minore”, quale l' “esclusivo riferimento all'interesse morale e materiale della prole” contenuto nell'art. 337 ter cpv. c. c., che ricalca fedelmente la formula di cui al testo originario dell'art. 6 L. 898 del 1970 (legge div.), utilizzata quindi in ambito municipale già in un tempo certamente risalente rispetto alle Carte ed alle Convenzioni sovranazionali appena ricordate.

ovvero ancora di mera icona legislativa (per fermarsi alle opinioni più accreditate) né dire, nella sostanza, in cosa tale interesse consista⁶.

Invece, può darsi qui per definitivamente acquisita la consapevolezza che l'interesse del minore vada determinato caso per caso e che lo stesso, per quanto preminente, non si sottragga affatto alla nota esigenza del bilanciamento⁷.

2. Il regolamento UE sulla I. A. e la tutela del minore

Il recente regolamento UE 1689/2024 sulla I. A. rivela sicura sensibilità per le possibili interferenze del fenomeno in questione con la posizione del minore e la sua tutela⁸.

Ricorrono tuttavia a tal ultimo riguardo previsioni prive di un accettabile grado di specificità e di spiccato carattere precettivo⁹; varrà al riguardo ricordare in particolare che:

- il considerando n. 29 dà atto che i sistemi di IA possono sfruttare le vulnerabilità di una persona o di uno specifico gruppo di persone dovute all'età;

- il considerando n. 48 riconosce che la portata dell'impatto negativo del sistema di I. A. sui diritti fondamentali protetti dalla Carta è di particolare rilevanza ai fini della classificazione di un sistema di IA tra quelli ad alto rischio, sottolineando la circostanza che i minori godono di diritti specifici sanciti dall'articolo 24 della Carta e dalla Convenzione delle Nazioni Unite sui diritti dell'infanzia e dell'adolescenza, ulteriormente sviluppati nell'osservazione generale n. 25 della Convenzione delle Nazioni Unite dell'infanzia e dell'adolescenza per quanto riguarda l'ambiente digitale, che prevedono la necessità di tenere conto delle loro vulnerabilità e di fornire la protezione e l'assistenza necessarie al loro benessere;

- ed infine, i considerando nn. 32 e 54 e l'art. 5 par. 1 lett. b) danno particolare rilievo alle discriminazioni in relazione a situazioni di specifica vulnerabilità che, tra l'altro, si evidenziano possano derivare anche dall'età.

Alla luce della rilevazione che precede appare conseguente ritenere che le modalità di tutela del minore anche nel contesto che ci occupa comporteranno, almeno allo stato dell'arte

⁶ La bibliografia in argomento è molto vasta ed il tema è affrontato, *ex professo* o incidentalmente, da tutti gli AA. che si occupano di diritto minorile. Può solo ricordarsi qui che dedica ai problemi indicati nel testo un contributo monografico M. DI MASI, *L'interesse del minore. Il principio e la clausola generale*, Napoli, 2020, *passim*; per un'analisi critica recente del concetto di interesse del minore cfr. L. Lenti, *Diritto della famiglia*, Milano, 2021, 73 ss. ed E. LAMARQUE, *Pesare le parole. Il principio del best interests of the child come principio del migliore interesse del minore*, in *Fam. dir.*, 2023, 365 ss.

⁷ Su questi profili è d'obbligo il rinvio a V. Scalisi, *Il superiore interesse del minore ovvero il fatto come diritto*, in *Riv. dir. civ.*, 2018, 405 ss.

⁸ Il tema oggetto di specifica attenzione in questo studio si iscrive ovviamente nella più generale problematica attinente l'idoneità del Reg. UE sull' I. A. a tutelare efficacemente le libertà fondamentali. I primi commenti non appaiono univocamente orientati: si vedano al riguardo le diverse prospettive offerte da G. VETTORI, *La "crisi del diritto" e il regolamento I.A. Un esempio di positivismo inclusivo*, in *Foro. it.*, 2025, V, 12 ss. e da S. PAGLIANTINI, *La base giuridica dell' AI Act ex art. 17ue: l'intelligenza artificiale tra mercato e persona*, *ivi*, 19 ss.

⁹ Il rilievo svolto nel testo vale naturalmente per gli enunciati riferiti direttamente al rilievo dell'età ed ovviamente non pregiudica la constatazione che il reg. U. E. sull'I. A. presenti, nel suo complesso, un approccio di tipo precettivo-sanzionatorio basato sul livello di rischio, nel senso che *a priori* vengono individuate, attraverso regole di carattere generale, pratiche a rischio *cd. inaccettabile*, che, come tali, sono assolutamente vietate in ragione proprio del loro impatto sui diritti fondamentali e sui valori della libertà, dignità, uguaglianza e democrazia e poi attività *ad alto rischio* in tesi lecite, ma accompagnate da obblighi stringenti per chi le attua.

e salvo la emissione di specifiche linee guida e di normative di dettaglio¹⁰, il mero adattamento dei presidi previsti in generale per realizzare tale protezione ovvero la estensione di tutele previste per ambiti contermini che sollevano questioni analoghe perchè strettamente collegate alla innovazione tecnologica¹¹.

Inalterate rimangono d'altro canto, pure nell'ambito oggi in esame, le esigenze di individuazione in concreto della consistenza e dei contorni dell'interesse del minore a cui dovranno concorrere - come di consueto - gli strumenti di auto ed eterodeterminazione previsti dalla normativa di settore.

3. *La deriva procedimentale/processualistica: la matrice prevalentemente giudiziale della determinazione dell'interesse del minore*

Proprio a quest'ultimo riguardo la rilevata ed ormai pacifica esigenza dell'effettività dei rimedi posti a servizio della tutela del minore ha legittimato l'osservazione che attraverso l'intervento legislativo posto in essere prima dalla Legge delega 206 del 2021 e quindi dal D. Lgs. 149 del 2022 (meglio noti come Riforma Cartabia), si sarebbe scardinata “la tradizionale

¹⁰ In ambito nazionale va rilevata l'approvazione da parte del Senato della Repubblica, nella seduta del 20.3.2025, di un disegno di legge di iniziativa governativa che contiene disposizioni e deleghe al Governo in materia di intelligenza artificiale e tra queste alcune norme a vario titolo riferibili ai soggetti minori di età. In particolare il comma 4 dell'art. 4 stabilisce che “L'accesso alle tecnologie di intelligenza artificiale dei minori di anni quattordici richiede il consenso di chi esercita la responsabilità genitoriale. Il minore di anni diciotto, che abbia compiuto quattordici anni, può esprimere il proprio consenso per il trattamento dei dati personali connessi all'utilizzo di sistemi di intelligenza artificiale, purché le informazioni e le comunicazioni di cui al comma 3 siano facilmente accessibili e comprensibili”; avuto riguardo al rapporto di lavoro si prevede che “L'intelligenza artificiale nell'organizzazione e nella gestione del rapporto di lavoro garantisce l'osservanza dei diritti inviolabili del lavoratore senza discriminazioni in funzione del sesso, dell'età, delle origini etniche, del credo religioso, dell'orientamento sessuale, delle opinioni politiche e delle condizioni personali, sociali ed economiche, in conformità con il diritto dell'Unione europea”; ed infine, a proposito della istruzione e della formazione, nell'art. 24 lett. e) e g), in vista dell'adozione della legislazione delegata, si fissano i principi della “previsione di percorsi di alfabetizzazione e formazione in materia di utilizzo dei sistemi di intelligenza artificiale” e del “potenziamento, all'interno dei curricula scolastici, dello sviluppo di competenze scientifiche, tecnologiche, ingegneristiche e matematiche legate alle discipline STEM, nonché artistiche, al fine di promuovere la scelta da parte delle studentesse e degli studenti, anche attraverso mirate attività di orientamento personalizzato, di percorsi di formazione superiore relativi alle menzionate discipline”. Lo *ius positum* conosce poi una copiosa legislazione, sia municipale che sovranazionale, che riguarda sotto diversi profili la tutela dell'interesse del minore nell'ambiente digitale. Un'ampia rassegna critica dello stato dell'arte si deve a B. Agostinelli, *Diritti e doveri del figlio*, sub art. 316 c. c., in *Commentario del Codice Civile Scialoja Branca Galgano*, a cura di G. De Nova, Bologna, 2024, 117 ss.; particolare attenzione alla normativa vigente nella materia viene dedicata pure da I. Garaci, *Il <<superiore interesse del minore>> nel quadro di uno sviluppo sostenibile dell'ambiente digitale*, in *Nuove leggi civ. comm.*, 2021, 800 ss.

¹¹ Intendiamo riferirci, come è ovvio, alla materia della tutela dei minori e più in generale delle persone vulnerabili dettata in tema di *privacy* e di trattamento dei dati. Su questo tema la letteratura è particolarmente estesa: per un panorama recente si può consultare, *Minori e privacy. La tutela dei dati personali dei bambini e degli adolescenti alla luce del regolamento (UE) 2016/679*, a cura di Alessandra Annoni e Arianna Thiene, *Atti del convegno* di Rovigo 22.11.2018, Napoli, 2019, *passim*, A. La Spina, *L'identità del minore nella realtà on-life tra protezione e autodeterminazione*, in *Fam. dir.*, 2024, 920 ss. e G. Cosco, *Protezione e autodeterminazione del minore nell'ambiente digitale*, in corso di pubblicazione in *Questa rivista*, che ho letto per gentile concessione dell'A.

dicotomia fra la dogmatica dei diritti e dei rimedi” inaugurandosi “un rinnovato approccio che riconosce l’inscindibile legame tra situazioni giuridiche sostanziali e tecniche di tutela processuale”¹².

Ma è appena il caso di osservare come l’auspicata effettività del rimedio sia inscindibilmente legata ad un funzionamento e ad una gestione efficienti della giustizia civile e penale e che nel nostro paese tale preconditione sia allo stato ben lungi dal potersi considerare esistente ed attuata.

Su questo giudizio negativo incidono in modo determinante i tempi del processo che, proprio nella materia della giustizia minorile in Italia, hanno condotto, come noto, a decise prese di posizione del Giudice sovranazionale, ma, allo stesso tempo, alla introduzione di previsioni normative che, almeno nelle intenzioni, sono dirette proprio al fine di accelerare i procedimenti giudiziari in materia minorile¹³.

Nonostante queste criticità non è dubbio che gli avanzamenti ordinamentali recenti in materia di tutela del minore siano decisamente orientati a privilegiare la determinazione dell’interesse del minore, non meno che il suo perseguimento, attraverso strumenti di matrice giudiziale.

In questo senso depone, tra l’altro, la traslazione nel codice di rito della disciplina dell’ascolto del minore, circostanza questa che, tuttavia, non può fare certo dell’istituto un mezzo di prova¹⁴, neppure atipico¹⁵, nè fare dimenticare che l’ascolto del minore è in primo luogo un diritto di quest’ultimo immediatamente riconducibile all’art. 21 della Costituzione¹⁶ e costituisce appunto strumento di manifestazione del pensiero e delle opinioni del minore, che deve trovare attuazione innanzitutto nella famiglia, costituendo, come è stato da più parti osservato, uno dei momenti caratterizzanti la partecipazione del minore alla vita della comunità familiare, al fine di rendere concretamente operante la funzione educativa¹⁷.

¹² G. Carapezza Figlia, *Effettività della tutela del minore e misure di coercizione indiretta. Gli artt. 614-bis e 709-ter c.p.c. nella riforma del processo della famiglia*, in *Dir. fam. pers.*, 2022, p. 632 ss. e, nel medesimo senso, *Idem*, *Prospettiva “paidocentrica” e attuazione dei doveri genitoriali nella nuova giustizia familiare*, in *Fam. dir.*, 2024, 103 ss, spec. 104 ed ivi note 10 e 11.

¹³ E’ quello indicato nel testo profilo di rilevanza centrale che riguarda la capacità dello strumento processuale e del sistema giudiziario di fornire una risposta urgente, quale è quella certamente richiesta nella materia minorile. Su questo punto ha espressamente disposto il Reg. U. E. 1111 del 2019, relativo alla competenza, al riconoscimento e all’esecuzione delle decisioni in materia matrimoniale e in materia di responsabilità genitoriale, e alla sottrazione internazionale di minori, con l’art. 24, rubricato “Celerità del procedimento giudiziario”. La posizione della Corte EDU sulla (in)idoneità dell’ordinamento italiano a fornire una tutela urgente all’interesse superiore del minore è espressa nella recente sentenza del 19 ottobre 2023 - Ricorso n. 48618/22 - Causa A.S. e M.S. c. Italia. Nella riforma Cartabia, certamente ispirata, nelle intenzioni, a garantire un’adeguata tutela del minore, si vedano in particolare gli artt. 473 bis 22 c. p. c. a proposito dell’adozione da parte del Giudice di provvedimenti “temporanei ed urgenti” e 473bis 6 ove si prevede, nel caso del rifiuto del minore di incontrare il genitore, che il Giudice proceda all’ascolto del minore “senza ritardo”.

¹⁴ Per la netta affermazione che “l’ascolto del minore non è un atto istruttorio, ma un diritto, esercitato dal minore capace di discernimento, di esprimere liberamente la propria opinione in merito a tutte le questioni e procedure che lo riguardano, vale a dire sulle questioni che hanno incidenza sulla sua vita e sulla relazione familiare” v. da ultimo Cass. 21.2.2025 n. 4595.

¹⁵ Questa la definizione prospettata da A. Graziosi, *Note sul diritto del minore ad essere ascoltato nel processo*, in *Riv. trim. dir. proc. civ.*, 1992, 1305.

¹⁶ Per questo rilievo sia permesso il rinvio a R. Amagliani, *L’interesse superiore del minore e la riforma Cartabia: profili sostanziali*, in *Dir. fam. pers.*, 2025, 131 ss, spec. 136 ss.

¹⁷ È questa la felice risalente intuizione, ampiamente argomentata sulla base dei principi costituzionali e collocata in una prospettiva sistematica, di F. Ruscello, *Garanzie fondamentali della persona e ascolto del minore*, in *Famiglia*, 2002, 933 ss; ma per la stessa opinione v. pure C. M. Bianca, *Il diritto del minore all’ascolto*, in *Nuove leggi civ. comm.*, 2013, 546, G. Ferrando, *La nuova legge sulla filiazione: profili sostanziali*, in *Corr. giur.*, 2013, 529 ed ora

Né può essere dubbio, per altro verso, che la valorizzazione e l'accentuata rilevanza degli strumenti processuali in vista della tutela del minore metta in luce le dinamiche familiari cogliendole essenzialmente nel momento patologico e quindi finendo per inevitabilmente trascurare gli argomenti che precedono i quali trovano collocazione elettiva in un contesto fisiologico delle relazioni tra genitori e figli.

Una prospettiva siffatta, proprio con riferimento al tema di cui qui si discute, finisce per ulteriormente radicalizzare la consueta e scontata contrapposizione, costantemente segnalata dagli studiosi dell'intelligenza artificiale¹⁸, ma anche a livello delle istituzioni sovranazionali¹⁹, tra rischi *versus* opportunità. Di tal chè è stato di recente affermato, a proposito della interferenza dei sistemi di A. I. nei confronti dei diritti della persona, che, se generalmente si è portati ad evidenziare vantaggi ed opportunità, sono i profili di rischio quelli sui quali occorre soffermarsi di più²⁰.

4. *La insoddisfacente prospettiva della contrapposizione rischi/opportunità per adeguatamente contestualizzare la posizione (e le prospettive di realizzazione dell'interesse) del minore nell'era dell'intelligenza artificiale.*

Le premesse operate, che in estrema sintesi vedono contrapposta una disciplina caratterizzata, almeno allo stato, da un alto grado di genericità e lo spostamento, secondo un indirizzo che pare ormai consolidato, del baricentro della tutela dal diritto sostanziale a quello processuale, conducono ad una conseguenza obbligata, e cioè a quella che è stata definita come “innaturale”²¹ ma necessaria supplenza giurisdizionale.

Ed invece l'avvento della I. A. può e deve servire alla scienza giuridica innanzitutto ad affinare (riempiendola di nuovi e più aggiornati contenuti) quella categoria, che ancora non

anche L. QUERZOLA, *L'ascolto del minore nel processo civile, tra diritto di libertà, mezzo di istruzione e strumento di partecipazione*, in *Rivista trimestrale di diritto processuale e civile*, 2023, 1341.

¹⁸ Con particolare incisività, sul punto, cfr. R. PARDOLESI, *Intelligenza artificiale (de)generativa? Svolazzi a cavallo tra ovvietà e disincanto*, in *Foro it*, 2025, V, 5 ss.

¹⁹ Sul punto, in una prospettiva generale, cfr. la relazione dell'Agenzia dell'Unione Europea per i diritti fondamentali, *Preparare un giusto futuro. L'intelligenza artificiale e i diritti fondamentali*, Lussemburgo, 2021, *passim*.

²⁰ E. BATTELLI, *Necessità di un umanesimo tecnologico: sistemi di intelligenza artificiale e diritti della persona*, in *Diritto di famiglia e delle persone*, 2022, 1107.

²¹ In questi termini A. RUGGERI, *La Costituzione, le sue trasformazioni, la sua scienza (profili metodico-teorici)*, a cura di A. MORELLI e A. RUGGERI, Napoli, 2024, 91 ss.

ha forse ricevuto una soddisfacente sistemazione²², costituita dalla capacità di discernimento, che si contrappone alla capacità di agire²³ ed è destinata a sostituire quest'ultima nell'ambito delle attività del minore che costituiscono l'esplicazione più rilevante della sua capacità di autodeterminazione, e che quindi vanno distinte dagli atti ad esclusivo contenuto e rilievo patrimoniale²⁴.

²² Sul punto la dottrina non è pervenuta ancora a risultati definitivi in ragione di un materiale normativo estremamente eterogeneo, ma anche in virtù di una sorta di “ineffabilità” propria della figura in questione (per questa terminologia v. B. AGOSTINELLI, *op. cit.*, *sub comm. art. 315 bis c. c.*, 90). Invero, se è comunemente segnalata la inidoneità e la inadeguatezza dell'istituto della capacità di agire, legata al raggiungimento della maggiore età, a giustificare una serie numerosa di ipotesi, talune oggetto di espressa previsione normativa, nelle quali si riconosce al minore la possibilità di compiere validamente atti rilevanti *sub specie iuris* (da ultimo, per questo rilievo, I. GARACI, *op. cit.*, 804), va tuttavia nello stesso tempo rilevato che di codesti atti non può denotarsi una connotazione unitaria. A tal uopo non appare funzionale ed esaustiva la contrapposizione - di cui si dà conto nel testo - tra atti a contenuto patrimoniale, che rimarrebbero sotto il governo del tradizionale istituto della capacità di agire e della disciplina che la riguarda, e atti di natura personale o personalissima che, invece, dovrebbero essere valutati alla luce della capacità di discernimento. Varrà al riguardo sottolineare, infatti, che quest'ultima serve, innanzitutto, a fare salva la validità degli atti correnti della vita quotidiana di schietto contenuto patrimoniale compiuti non solo dal minore di età, ma pure dal beneficiario dell'amministrazione di sostegno (ai sensi dell'art. 409 c. c.) e dall'interdetto e dall'inabilitato (ai sensi dell'art. 427 primo comma c. c.); per questo rilievo cfr. A. Sassi, *La genitorialità nel sistema di protezione della persona*, in *Riv. trim. dir. proc. civ.*, 2025, 25 e nello stesso senso I. Garaci, *op. cit.*, 805. Ed ancora. L'istituto della capacità di discernimento secondo la *communis opinio* viene ritenuto preferibile a quello della capacità di agire perchè più duttile e meno legato al parametro rigido del raggiungimento di una età determinata (18 anni, allo stato). Non può negarsi, però, che l'operatività delle previsioni dettate in materia di capacità di discernimento sia anch'essa in certa misura condizionata dal requisito dell'età, nel senso che in talune ipotesi, ove il minore abbia conseguito una età determinata, la capacità di discernimento si presume, così come avviene, solo per esemplificare, nell'art. 315 bis c. c. a proposito del diritto all'ascolto del figlio dodicenne (per questa opinione, v. B. Agostinelli, *op. ult. cit.*, 91; ma nello stesso senso cfr. Cass. 21.2.2025 n. 4595, cit.), mentre in altre non rileva affatto. Ciò avviene ad esempio nell'art. 250 c. c., in tema di assenso del figlio quattordicenne nei confronti del riconoscimento del genitore naturale, che costituisce, secondo A. Scalera, *sub comm. art. 250 c. c.*, in *Commentario del codice civile*, diretto da E. Gabrielli, *Della famiglia*, a cura di G. Di Rosa, Milano, 2018, 181, l'esercizio da parte del minore di una libertà insindacabile espressione di una capacità speciale, ma altrettanto si può dire avuto riguardo alla cd. maturità digitale sulla quale dispongono l'art. 8 del GDPR e l'art. 2 quinquies inserito nel d. lgs. 193/1996 dal d. lgs. 101/2018, che ha fissato in quattordici anni l'età del minore “per esprimere il consenso al trattamento dei propri dati personali in relazione all'offerta diretta dei servizi della società dell'informazione”; per una valutazione critica della disciplina vigente in tale materia v. B. Agostinelli, *op. cit.*, *sub comm. art. 316 c. c.*, 117 ss. Infine, ove si consideri l'ambito della sfera dei diritti della personalità, e segnatamente il diritto alla salute del minore, ci si avvede come la L. 22 dicembre 2017 n. 219, pur avendo valorizzato, all'art. 3, le “capacità di comprensione e di decisione del minore” e sancito il diritto di quest'ultimo a “ricevere informazioni sulle scelte relative alla propria salute in modo consono alle sue capacità” (presupponendo quindi, pur senza dirlo esplicitamente, il rilievo della capacità di discernimento) finisce poi per consegnare senz'altro il consenso nelle mani dei genitori (per questi rilievi si veda B. Agostinelli, *op. cit.*, 130 ss.). Da questa sintetica ed incompleta rassegna si desume che la natura giuridica della capacità di discernimento non sia, almeno allo stato, suscettibile di ricostruzione unitaria.

²³ E che per taluno dovrebbe invece coincidere con la capacità di intendere e di volere: in questo senso appaiono orientati A. Sassi, *op. cit.*, 24, e G. Palmeri, *L'impatto della riforma nel sistema dei rapporti familiari in particolare il minore e il curatore speciale: la nuova normativa*, in *Dir. fam. pers.*, 2025, 291.

²⁴ Ove si intenda aderire, con riferimento alla posizione del minore, alla dicotomia tra diritti e profili personali e diritti e profili patrimoniali, “tralatizia, sebbene non sempre agile” secondo B. Agostinelli, *op. cit.*, 115.

La ricostruzione dogmatica dell'istituto, ammesso che sia possibile ed utile proporla la ricerca²⁵, non può essere qui nemmeno tentata²⁶, anche se di fondamentale rilievo appare a questo fine una corretta individuazione dell'istituto dell'ascolto del minore²⁷.

Si deve tuttavia subito constatare ed ammettere che i sistemi di I. A. hanno una fenomenale idoneità ad incidere sul bagaglio delle informazioni a disposizione del minore ed a fornirgli una serie potenzialmente infinita di suggestioni (gli *input* di cui appunto ci parla la ricordata definizione del Reg. U. E.) in grado di incidere in modo assai significativo sull'acquisizione da parte del minore stesso della maturità di pensiero e quindi sulla sua capacità di discernimento, ed in definitiva di avere un peso determinante sulla formazione della sua personalità.

Le esemplificazioni sono agevoli e numerose: si va dagli algoritmi che sulle piattaforme *streaming* (*disney plus*, ad esempio) forniscono ai bambini consigli su quali video guardare, agli *smart speaker* che si utilizzano per ascoltare musica, recuperare informazioni e creare promemoria; dagli altoparlanti intelligenti (si pensi ad Alexa) oggi assai diffusi nelle famiglie ai *software* di apprendimento adattativo (cioè basati sulle peculiarità del singolo studente).

L'obiezione è ovviamente dietro l'angolo: chi assicura, si dirà, la correttezza delle informazioni e l'adeguatezza degli stimoli (dei *feedback*, secondo una terminologia ormai *a la page*).

Si tratta allora di riscoprire, ovvero di valorizzare adeguatamente, il ruolo delle formazioni sociali che, nella dimensione in esame, hanno ed assumono vieppiù un'importanza determinante, ed in primo luogo della famiglia e della scuola che devono riappropriarsi o, comunque, fare pienamente valere le competenze loro proprie, direi

²⁵ Lo nega P. STANZIONE, *Capacità, legittimazione, status*, in *Trattato di diritto civile e commerciale*, Cicu-Messineo, Milano, 2017, 131, rilevando che la capacità di discernimento deve intendersi «non già come nuova categoria dogmatica, bensì come valutazione casistica della situazione globale del minore in relazione al singolo atto, alla singola scelta esistenziale che lo stesso sia di volta in volta chiamato a compiere; come strumento per dare attuazione nel rispetto dell'essere e del divenire del minore a “quel principio di effettività” sancito con l'art. 3 cost.». Della stessa opinione è la giurisprudenza laddove afferma che “la capacità discernimento non è una nozione fissa ed è tendenzialmente ricollegata all'acquisizione di competenze intellettuali e concettuali che aiutino il minore a riconoscere e valutare razionalmente i dati provenienti al di fuori della propria dimensione personale. La stessa è dunque considerata sussistente in tutte le ipotesi in cui il minore sia in grado di cogliere dati, informazioni e stimoli provenienti dall'esterno, riguardanti la propria sfera esistenziale ed elaborarli secondo il proprio personale sentire, formandosi un proprio convincimento riguardo ad essi, le sue esigenze e i suoi bisogni”, così Cass. 7.3.2023 n. 6802 in *Fam. dir.*, 2023, 550 ss. con nota di F. TOMMASEO, *Una giurisprudenza evolutiva in tema di decisorietà e di audizione dell'infradodicesenne*, *ivi*, 554 ss.

²⁶ Si può qui solo rimandare agli AA. che, sotto vari aspetti, hanno contribuito a delineare i caratteri peculiari dell'istituto: tra gli altri, F. GIARDINA, *La maturità del minore nel diritto civile*, in *Nuova Giur. Civ. Comm., Giustizia minore? La tutela giurisdizionale dei minori e dei “giovani adulti”*, 2004, 95 ss. spec. 100 e *Idem*, *Interesse del minore: gli aspetti identitari*, *ivi*, 2016, 159 ss., P. STANZIONE, *op. cit.*, 131 ss., F. RUSCELLO, *Minore età e capacità di discernimento: quando i concetti assurgono a “supernorme”*, in *Fam. dir.*, 2011, 404 ss. Lo stato dell'arte si trova riassunto adesso in M. A. IANNICELLI, *Il diritto del minore ad essere ascoltato*, in *Nuove leggi civ. comm.*, 2023, 1065 ss.

²⁷ E ciò perchè, in larga misura, la capacità di discernimento pare essere considerata dal legislatore quale presupposto per procedere all'ascolto di minori al di sotto di una certa età, pur se il rapporto è esattamente contrario nella misura in cui soltanto l'ascolto può condurre all'apprezzamento della capacità di discernimento: in questo senso cfr. F. DANOVI, *Ascolto del minore, capacità di discernimento e obbligo di motivazione (fra presente e futuro)*, *Fam. dir.*, 2022, 997. Per una considerazione sistematica della disciplina in materia di diritto all'ascolto del minore cfr. M. TRIMARCHI, *Diritto di esprimere le proprie opinioni e ascolto del minore*, in *Fam. dir.*, 2024, 595 ss.

istituzionalmente, che vanno tuttavia esercitate in un contesto ordinamentale profondamente mutato rispetto a quello in cui certi modi di pensare si sono formati e quindi consolidati²⁸.

Il minore visto quale soggetto essenzialmente titolare di diritti è acquisizione ormai pacifica anche se relativamente recente²⁹ e vede come contraltare (in luogo della potestà dei genitori) l'affermazione della responsabilità genitoriale³⁰.

Altrettanto deve dirsi della configurazione della istruzione come oggetto di una specifica situazione giuridica soggettiva attiva del minore medesimo da far valere su un duplice fronte: quello privato, verso la famiglia, quello pubblico, nei confronti delle istituzioni³¹.

Fatta questa premessa la contrapposizione, onnipresente nel discorso sul diritto minorile, tra una prospettiva adultocentrica ed un'altra puerocentrica finisce per diventare oziosa e puramente nominalistica, dovendo invece mettersi in risalto che le attività di cura e di assistenza morale, così come quelle volte alla formazione costituiscono oggetto di doveri formalizzati già a livello costituzionale (artt. 30 e 34 Cost.) e non possono che avere di mira in via esclusiva lo sviluppo della personalità del minore.

Sicché l'avvento e la diffusione dell'intelligenza artificiale in tutti gli aspetti della vita odierna impongono innanzitutto ai genitori ed ai soggetti deputati alla formazione di assumere in primo luogo adeguata consapevolezza del fenomeno (evitando accuratamente sia anacronistiche prese di distanza così come affidamenti fideistici) e quindi rendono doverosi i comportamenti coerenti alla imprescindibile indicazione costituzionale contenuta nell'art. 2.

A questo fine alcuni punti possono essere fermati:

- Deve tenersi conto che la personalità del minore, così come la capacità di discernimento, è strettamente legata alla sua maturazione psicofisica che non costituisce un dato né astratto né costante, ma ammette variabili e chiede l'accertamento in concreto attraverso lo strumento dell'ascolto³².

- Deve preservarsi quello che è stato chiamato umanesimo tecnologico³³, e cioè un approccio alla tecnologia che abbia sempre presente la centralità della persona e della sua tutela.

- E con particolare riferimento alla situazione del minore, occorre avere ben chiaro che l'intelligenza artificiale non può giammai sostituire le indispensabili ed imprescindibili relazioni personali tra il minore ed i componenti delle formazioni sociali a cui quest'ultimo a vario titolo partecipa, in primo luogo, ancora una volta, la famiglia e la scuola³⁴.

²⁸ Su questi aspetti, fondamentali per affrontare adeguatamente il tema in esame, v. I. GARACI, *op. cit.*, 811 ss. e N. MACCABIANI, *Protezione ed empowerment dei minori nell'era dell'intelligenza artificiale: coordinate costituzionali*, *Rivista di biodiritto*, 2024, 347 ss, spec. 355 ss.

²⁹ Lo stato dell'arte e l'evoluzione della disciplina si trovano delineati in M. Bianca, *Minore (tutela sostanziale in generale)*, in *Enc. dir., I Tematici, Famiglia*, Milano, 2022, 886 ss.

³⁰ Il percorso che ha condotto dalla potestà dei genitori all'affermazione della responsabilità genitoriale è tracciato in A. DI LANDRO, *Responsabilità genitoriale verso i figli*, in *Enc. dir., I Tematici, Famiglia*, cit., 1161 ss., spec. 1181 ss.

³¹ *Ibidem*, 1169.

³² Si vedano soprattutto i riferimenti contenuti *supra* nelle note nn. 23, 24 e 25.

³³ Per questa terminologia cfr. E. BATTELLI, *Necessità di un umanesimo tecnologico: sistemi di intelligenza artificiale e diritti della persona*, cit., spec. 1118 ss.

³⁴ Su questo aspetto si insiste particolarmente in Antiqua e nova. *Nota sul rapporto tra intelligenza artificiale e intelligenza umana*, a cura del Dicastero per la Dottrina della Fede e del Dicastero per la Cultura e l'Educazione, Milano, 2025, spec. 50 ss. Su implicazioni specifiche della ineludibile esigenza segnalata nel testo cfr., tra gli

5. Il caso *Replika*

Per dare concretezza al discorso sin qui condotto un cenno finale merita il caso “Replika”.

Si tratta di una *chatbot* dotata di interfaccia scritta e vocale che, basandosi sull’intelligenza artificiale, è in grado di generare un “amico virtuale”.

Replika, secondo i presentatori, è sempre disponibile ad ascoltare problemi, consigliare e confortare ed è in grado di agevolare il benessere emotivo dell’utente consentendo a chiunque interagisce con lui di ridurre l’ansia, gestire lo stress ed agevolare la socializzazione.

Ebbene da parte di alcuni utenti è stato contestato il fatto che l’applicazione risultava nei confronti dei minori troppo invadente o anche molesta.

Della questione si è interessato il Garante della privacy che con provvedimento d’urgenza del 2.2.2023³⁵ ha bloccato l’utilizzazione in Italia di Replika rilevando, per quanto qui interessa, che «l’applicazione presenta concreti rischi per i minori di età, a partire dalla proposizione ad essi di risposte assolutamente inidonee al loro grado di sviluppo» essendo sprovvista di qualsiasi meccanismo di verifica dell’età.

Forse occorrerebbe almeno domandarsi se prima ancora di indagare sulla possibilità dell’applicazione di I. A. di discernere tra maggiorenne o minore, o tra minore bambino, adolescente o quasi maggiorenne, non si mostri necessario verificare come venisse esercitata nei casi segnalati al Garante la funzione educativa e di formazione da parte dei soggetti a ciò deputati.

Ed ovviamente concludere che in ben altra dimensione si pongono sistemi di I. A., pur del tutto simili a quello in esame, allorquando siano predisposti, come avviene in alcuni reparti di pediatria oncologica di ospedali americani, per offrire quella che è stata definita l’informatica affettiva ai piccoli pazienti in isolamento totale per motivi terapeutici³⁶.

altri, G. DI ROSA, *I robot medici*, in *Pers. merc.*, 2022, spec. 16 ss. e S. ATTOLLINO, *Relazioni umane, fede e tecnologie. Il nuovo volto della famiglia digitale*, *Dir. fam. pers.*, 2024, 242 ss.

³⁵ Si tratta del provvedimento n. 39 del 2.2.2023 prot. n. 18321/23 che, nella versione integrale, si trova sul sito dell’Autorità Garante per la Protezione dei Dati Personali. Il Garante, più di recente, è intervenuto nuovamente sulla questione con il provvedimento n. 232 del 10.4.2025, con il quale è stata sanzionata la soc. Luka Inc. per non avere osservato le disposizioni del GDPR in tema di implementazione dei dati nella *chatbot* con l’applicazione di una sanzione di cinque milioni di euro.

³⁶ Si tratta “del robot umanoide *Pepper* (traduzione di pepe, peperoncino, facendo riferimento, per traslato, al brio e alla vitalità), che presenta tratti di significativa interattività; esso è infatti in grado di muoversi nell’ambiente circostante (spostandosi sulle rotelle), riconoscere un viso umano e percepire (tramite appositi sensori) le emozioni delle persone con cui entra in relazione. In particolare, in alcune strutture sanitarie tale robot è stato destinato, all’interno dei reparti pediatrici, a supporto delle figure professionali quale facilitatore della relazione tra la figura del medico (o di altro sanitario) e i piccoli pazienti. Si tratta, pertanto, di uno strumento deputato a migliorare la qualità delle cure e ridurre il disagio della permanenza dei piccoli pazienti nella struttura di degenza” secondo quanto riferisce, nei termini di cui sopra, G. DI ROSA, *Profili giuridici della robotica medica, in diritto e tecnologia. Precedenti storici e problematiche attuali*, *Atti delle giornate di studi* (Catania, 8 ottobre 2021 - 21 e 22 ottobre 2022 25 novembre 2022 - 19 e 20 maggio 2023) a cura di Giovanni Di Rosa, Sara Longo e Tommaso Mauceri, 90.



FRANCESCO CIRAIOLO*

CREDIT SCORING ALGORITMICO E TUTELA DELLA PERSONA.

RIFLESSIONI INTORNO AI CASI *SCHUFA* (CGUE, C-634/21) E *DeB* (CGUE, C-203/22)

SOMMARIO: 1. Attività di valutazione del merito creditizio e progresso tecnologico. Il *credit scoring* algoritmico. - 2. Il *credit scoring* algoritmico tra luci e ombre. - 3. Il *credit scoring* algoritmico come processo decisionale automatizzato (art. 22 GDPR). Il caso *Schufa* (CGUE, 7 dicembre 2023, C-634/21). - 4. (Segue). La tutela della persona nel processo decisionale automatizzato. Il diritto ad ottenere «*significant information*» sulla logica della decisione algoritmica. - 5. (Segue). Diritto alla spiegazione della decisione automatizzata e *full disclosure* dell'algoritmo. Il caso *DeB* (CGUE, 27 dicembre 2025, C-203/22). - 6. Le ulteriori affermazioni normative del diritto alla spiegazione della decisione algoritmica. La direttiva UE 2023/2225 sul credito ai consumatori (CCD II) e il Regolamento UE 2024/1689 sull'intelligenza artificiale (*AI Act*). - 7. *Credit scoring* e *social scoring*. La ricerca di una linea di confine. - 8. Considerazioni finali. La trasparenza algoritmica, tra tecniche di *eXplainable AI* e strumenti di *enforcement*.

1. Attività di valutazione del merito creditizio e progresso tecnologico. Il *credit scoring* algoritmico

Le attività di valutazione del merito creditizio – intese, genericamente, come quelle attività volte a determinare le prospettive di adempimento, da parte del debitore, delle obbligazioni nascenti da un contratto di finanziamento – sembrano connotate da un evidente paradosso. E invero, sebbene il giudizio sulla solvibilità dei soggetti sovvenuti costituisca un passaggio imprescindibile nelle procedure di erogazione del credito, il quadro normativo di riferimento – seppur arricchitosi, nel tempo, di disposizioni via via più puntuali, volte a circoscrivere gli ampi margini di discrezionalità tradizionalmente riservati ai finanziatori - appare ancora frastagliato e incompleto, a partire, a ben vedere, dalla mancanza di una definizione univoca della nozione stessa di “merito creditizio” (con incertezze che riguardano, di conseguenza, anche le variabili da considerare nel contesto della relativa valutazione).

La lacunosità del quadro regolatorio è problema nient'affatto secondario, se solo si osserva che il giudizio di affidabilità creditizia di coloro che richiedono un prestito rappresenta un adempimento posto a garanzia di interessi generali, che esulano, cioè, dalla protezione delle singole parti del contratto (l'una, il creditore, interessata ad ottenere la restituzione, nei tempi previsti dal piano di ammortamento, del capitale e dei relativi interessi,

* Professore ordinario di Diritto dell'economia, Università degli Studi di Messina.

in un'ottica di conformità al canone normativo della sana e prudente gestione¹; l'altra, il debitore, interessata ad ottenere un finanziamento adeguato ai propri bisogni ed economicamente sostenibile, ossia commisurato alle proprie effettive capacità restitutorie), per estendersi, più significativamente, alla tutela del buon funzionamento e della stabilità dell'intero sistema bancario. La valutazione del merito creditizio, in altri termini, è posta a presidio di un efficiente processo di selezione dei prenditori di credito e quindi, di riflesso, di una corretta allocazione delle risorse finanziarie, sicché impatta sulla creazione di un mercato del credito efficiente e stabile².

L'importanza delle attività in esame, insomma, richiederebbe la presenza di regole lineari ed esaustive, in grado di conferire alla relativa disciplina quel grado di certezza di cui la stessa necessita.

Ora, a fronte di tale esigenza, sul piano normativo si registra l'adozione di un primo nucleo di regole di matrice strettamente prudenziale, volte a disciplinare, segnatamente, il rischio di credito cui sono esposti i singoli intermediari (dove l'introduzione di disposizioni volte ad imporre a questi ultimi sofisticati modelli di valutazione delle proprie esposizioni³), cui si sono aggiunte, nel tempo, alcune norme settoriali, in materia di credito al consumo⁵, nel cui ambito la valutazione del merito creditizio, ad opera dei finanziatori, è espressamente concepita come strumento, di applicazione obbligatoria, necessario ad assicurare l'erogazione del c.d. prestito responsabile (potendosi quindi affermare che l'approccio del legislatore, in

¹ Art. 5 d. lgs. n. 385/2003 (*Testo unico bancario*, d'ora in poi TUB).

² Cfr., di recente, A. A. DOLMETTA, *La valutazione del merito del credito nell'accesso al servizio. La prospettiva del contratto di impresa*, in *Banca Borsa Titoli di Credito*, 2023, n. 3, pp. 320-1. In giurisprudenza, CGUE, 11 gennaio 2024, causa C-755/22; Cass., 30 giugno 2021 (ord.), n. 1861, ove si legge che l'obbligo di valutazione del merito creditizio «è posto, dal diritto positivo, ai fini della protezione dell'intero sistema economico dai rischi che una concessione imprudente o indiscriminata del credito bancario comporta».

³ Non è possibile intrattenersi in questa sede sulle diverse metodologie di *creditworthiness assessment* in atto esistenti, tradizionalmente articolate in *credit scoring* e *credit rating*. In arg., WORLD BANK, *Credit Scoring Approaches Guidelines*, 2019, pp. 3-4.

⁴ Si fa riferimento, in particolare, alle disposizioni sul contenimento delle varie tipologie di rischio (e del rischio di credito in particolare) contenute nelle norme di matrice europea (CRD/CRR), nonché nei provvedimenti adottati dalle autorità di vigilanza europee e nazionali (ABE, *Orientamenti in materia di concessione e monitoraggio dei prestiti*, 20 maggio 2020; BCE, *Linee Guida per le autorizzazioni delle c.d. FinTech Banks*, 2018; Banca d'Italia, *Disposizioni di vigilanza per le banche*, Parte I, Titolo IV, Cap. 3, All. A e *Disposizioni di vigilanza per gli intermediari finanziari*, Titolo III, Cap. I, Sez. VII). Per ulteriori dettagli si rimanda a F. CIRAIOLO, *L'attività di valutazione del merito creditizio tra lacune normative e mercati in trasformazione*, in AA. VV., *Studi in onore di Sabino Fortunato*, Bari, 2023, IV, p. 2747 ss.; E. CECCHINATO, *Note sulla disciplina della verifica del merito creditizio: per una sua rilettura alla luce della buona fede precontrattuale*, in *Rivista di diritto bancario*, 2023, n. 3, p. 457 ss.

⁵ Direttiva 2008/48/CE del Parlamento europeo e del Consiglio, del 23 aprile 2008, relativa ai contratti di credito ai consumatori (CCD), in GUUE L 133/66 del 22 maggio 2008, e direttiva 2014/17/UE del Parlamento europeo e del Consiglio, del 4 febbraio 2014, in merito ai contratti di credito ai consumatori relativi a beni immobili residenziali (MCD), in GUUE L 60/34 del 28 febbraio 2014, con relative disposizioni di attuazione. Sulla materia è destinata a incidere la recente direttiva UE 2023/2225 del Parlamento europeo e del Consiglio, del 18 ottobre 2023, relativa ai contratti di credito ai consumatori e che abroga la direttiva 2008/48/CE (CCD II), in GUUE L del 30 ottobre 2023 (sulla quale v. *infra*, par. 6). Per completezza, va segnalato che alle norme dell'ordinamento bancario italiano sulla valutazione del merito creditizio del consumatore (artt. 124-bis e 120-undecies TUB) fanno da complemento alcune disposizioni del nuovo Codice della crisi d'impresa e dell'insolvenza in punto di ristrutturazione dei debiti del consumatore (artt. 68 e 69 d. lgs. n. 14/2019), che prevedono specifiche conseguenze sanzionatorie a carico del finanziatore che non abbia correttamente espletato tali attività valutative (in arg., A. A. DOLMETTA, *"Merito del credito" e procedure di sovraindebitamento*, in *Fall.*, 2021, pp. 1214-5; M. FRANCHI, *Il ruolo del merito creditizio nella rinnovata disciplina in tema di composizione della crisi da sovraindebitamento: la chiusura di un cerchio?*, in *Rivista di diritto bancario*, 2021, pp. 527-8).

tale campo, mira a coniugare il buon funzionamento del mercato del credito con la protezione del debitore-consumatore dal rischio di sovraindebitamento)⁶.

Tali riferimenti normativi, tuttavia, non sono risultati idonei ad offrire adeguate risposte ai numerosi dubbi che le attività di *creditworthiness assessment* sono in grado di ingenerare. Sono note, infatti, le diatribe dottrinarie in ordine all'esistenza, perlomeno nell'ordinamento italiano, di un vero e proprio obbligo di astensione dalla concessione del finanziamento, in caso di giudizio negativo sulle condizioni del richiedente⁷, ovvero sulla natura dei rimedi (demolitori o risarcitori) esperibili dal debitore, in caso di omessa o errata valutazione del merito di credito, cui abbia fatto seguito l'erogazione di un prestito rivelatosi insostenibile⁸. Questioni che, nel contesto normativo ad oggi vigente, non sembrano trovare pacifiche soluzioni.

A tali rilievi si aggiunga che sulle attività in esame ha inciso in modo tutt'altro che indifferente il progresso tecnologico, che ha dato vita ad innovative modalità di apprezzamento del merito creditizio, basate su sofisticati sistemi di intelligenza artificiale, note come *algorithmic credit scoring*⁹. Queste ultime, tuttavia, non soltanto ripropongono le medesime problematiche giuridiche che caratterizzano ogni attività di valutazione del merito creditizio, ma ne prospettano di nuove e ulteriori, essenzialmente afferenti, per via delle peculiari modalità di svolgimento del servizio, a due macroaree di interesse: la disciplina dei

⁶ Con l'espressione "prestito responsabile" si intende fare riferimento ad un prestito che appaia, al contempo, profittevole per il finanziatore e "sostenibile" per il sovrvenuto. Il principio del prestito responsabile appare tuttavia connotato da una certa ambivalenza (F. TRAPANI, *Art. 124-bis*, in C. COSTA, A. MIRONE (a cura di), *Commento al testo unico delle leggi in materia bancaria e creditizia*, Torino, 2024, t. II, p. 2019 ss.; G. LIBERATI BUCCIANI, *Il merito creditizio del consumatore*, Milano, 2019, p. 20 ss.), potendo articolarsi nella duplice componente del *responsible borrowing* (ove si voglia porre l'accento sul profilo dell'autoresponsabilità del debitore rispetto alle proprie scelte di natura finanziaria) e del *responsible lending* (inteso come dovere del finanziatore di non concedere alla controparte prestiti non sostenibili, per prevenirne l'eccessivo indebitamento).

⁷ In senso affermativo, M. SEMERARO, *Informazioni adeguate e valutazione del merito creditizio: opzioni interpretative nel credito ai consumatori*, in *Rivista di diritto civile*, 2021, p. 710 ss.; T.V. RUSSO, *Valutazione ingannevole del merito creditizio e intervento perequativo del giudice nei contratti di credito immobiliare ai consumatori*, in *Contratto e impresa*, 2020, pp. 1500-1502; M. FRANCHI, *Il ruolo del merito creditizio*, cit., pp. 516-7, secondo il quale, tuttavia, il creditore potrebbe comunque dare corso al finanziamento, se il debitore, una volta informato della non sostenibilità del prestito, decidesse ugualmente di procedere. Da ultimo, C. A. VALENZA, *Obbligo di valutazione del merito creditizio del consumatore e regole in materia di banche dati*, in *Rivista trimestrale di diritto dell'economia*, 2024, suppl. al n. 4 a cura di P. GAGGERO, pp. 259-9, che evidenzia come un obbligo di astensione del creditore dalla concessione del finanziamento, in caso di valutazione negativa del merito creditizio, sia ricavabile con chiarezza dall'art. 18, par. 6, CCD II (norma che, peraltro, riproduce il contenuto dell'art. 18, par. 5, lett. a), MCD, non trasposto nell'ordinamento italiano in sede di attuazione della direttiva).

⁸ In letteratura sono state proposte soluzioni che spaziano dalla nullità del contratto per violazione di norme imperative (art. 1418 c.c.) o per difetto di causa concausa, all'annullabilità per dolo (ove il finanziatore abbia ingannato il debitore circa le sue effettive capacità di restituzione del prestito) o per errore (in caso di falsa rappresentazione che abbia inciso in modo determinante sulla volontà di stipulare il contratto di finanziamento), ma l'opinione più consolidata è quella che ravvisa una mera responsabilità precontrattuale dell'intermediario, per violazione di una regola di condotta riferibile alla fase precedente la stipula del finanziamento (per una sintesi delle diverse posizioni, si rimanda ancora a F. CIRAULO, *L'attività di valutazione del merito creditizio*, cit., p. 2753 ss.; F. TRAPANI, *Art. 124-bis*, cit., p. 2031 ss.; C. A. VALENZA, *Obbligo di valutazione del merito creditizio*, cit., p. 262 ss.).

⁹ Non a caso, in una recente pubblicazione di Banca d'Italia (A. LENTINI, D. E. MUNTEANU, F. ZENNARO, *Metodologia di classificazione del Fintech*, Roma, giugno 2025), il *credit scoring* è definito, in modo onnicomprensivo, come «Sistema automatizzato per valutare il merito creditizio delle controparti che si basa sull'applicazione di metodi o modelli statistici o AI/ML (Artificial Intelligence/Machine Learning)» (p. 16).

provvedimenti basati sul trattamento automatizzato di dati personali, da un lato, e la disciplina dell'intelligenza artificiale, dall'altro.

Su tali aspetti ci soffermeremo specificamente (tralasciando, invece, le questioni civilistiche cui si è dapprima fatto cenno), con il duplice obiettivo di individuare i rischi che le attività di valutazione automatizzata del merito creditizio presentano per la persona umana e le tutele ricavabili dall'attuale quadro regolatorio, tenuto conto anche dei provvedimenti legislativi di più recente adozione (CCD II e *AI Act*).

Nel condurre l'indagine ci gioveremo, in particolare, di alcune recenti pronunce della Corte di Giustizia, che offrono le prime, interessanti risposte alle complesse questioni sollevate dal *credit scoring* algoritmico, quale forma di processo decisionale in forma automatizzata. Il dato giurisprudenziale, letto unitamente a quello normativo, ci consegnerà una prospettiva sufficientemente ampia sul tema in esame, alla luce della quale potremo rassegnare le nostre conclusioni.

2. *Il credit scoring algoritmico tra luci e ombre*

La valutazione automatizzata del merito creditizio è una tecnica già da tempo utilizzata, sia pure in forme più embrionali rispetto a quelle affinate negli ultimi anni, da banche ed intermediari finanziari, che si avvalgono spesso, a tal fine, dei servizi offerti da società terze, specializzate nella produzione di *credit scoring*¹⁰.

Concettualmente, il ricorso alle più moderne tecnologie (*big data analytics* e *machine learning*), non muta l'essenza dell'attività di *creditworthiness assessment*, il cui obiettivo finale resta sempre quello di formulare un giudizio di solvibilità del richiedente il prestito, attraverso l'analisi di informazioni ritenute rilevanti. Ciò che è radicalmente cambiato, rispetto alle tradizionali forme di scrutinio della clientela bancaria, è invece la capacità di affidare le valutazioni di meritevolezza creditizia all'elaborazione di enormi quantità di dati, anche diversi da quelli, solitamente esaminati dagli intermediari, concernenti le condizioni reddituali, finanziarie e patrimoniali dell'interessato, o comunque ricavabili dall'andamento dei suoi pregressi rapporti con il finanziatore (c.d. *hard data*). La valutazione del livello di affidabilità di colui che intende accedere al credito, in altri termini, viene operata mediante algoritmi complessi, in grado di processare le più svariate ed eterogenee informazioni (c.d. *soft data*, ottenibili dalle interazioni sui *social media*, dalla navigazione in rete, dall'uso di *devices*

¹⁰ L'esempio più noto è costituito dal sistema statunitense *FICO Score*, il cui funzionamento si basa su dati raggruppati in cinque diverse categorie (storia creditizia del cliente, reddito, lunghezza della storia creditizia, tipologia dei finanziamenti già richiesti e finanziamento per il quale è presentata la nuova richiesta), ciascuna delle quali è munita di uno specifico peso ai fini della formulazione del punteggio rappresentativo del grado di solvibilità del soggetto valutato. Nuove e più sofisticate metodologie di *creditworthiness assessment* sono oggi utilizzate soprattutto da enti non autorizzati all'esercizio del credito (*start-up*, piattaforme digitali, società di informazione creditizia, ecc.), che forniscono tali servizi agli intermediari finanziari, in regime di *outsourcing* (G. L. GRECO, *Credit scoring 5.0, tra Artificial Intelligence Act e Testo Unico Bancario*, in *Rivista trimestrale di diritto dell'economia*, 2021, suppl. al 3, p. 74 ss.).

tecnologici, dalla rilevazione delle abitudini di spesa del consumatore, ecc.)¹¹, anche attraverso forme di autoapprendimento fondate sull'esperienza¹².

Appare ancora prematuro, allo stato dell'arte, pronunciarsi in merito all'efficienza e all'attendibilità di tali forme operative, non disponendosi ancora di dati empirici sufficientemente consolidati. Al tema si accompagna, dunque, un ampio grado di incertezza, derivante dalla constatazione che gli indubbi vantaggi offerti dall'uso di procedure di *algorithmic credit scoring* sembrano, in misura più o meno consistente, neutralizzati da rischi e criticità altrettanto significativi¹³.

Così, ad esempio, la possibilità di sfruttare, ai fini delle attività istruttorie, informazioni di natura eterogenea e assai diverse da quelle utilizzate nei sistemi tradizionali, consente sì di dare accesso al credito anche a soggetti *unbanked* (di solito non valutabili, in quanto privi di una pregressa storia creditizia), favorendo i processi di inclusione finanziaria (e dunque anche sociale) di soggetti ai margini del circuito finanziario, ma è contrastata dal rischio di un trattamento discriminatorio nei confronti di persone appartenenti a specifiche categorie (vittime di pregiudizi di tipo storico o statistico), determinato dai meccanismi decisionali propri dell'algoritmo¹⁴. Emerge invero, al riguardo, il problema di fondo che connota tutti i sistemi decisionali automatizzati, ossia il rischio che, per la presenza di *biases* nei dati utilizzati, per difetti di programmazione del *software* o per via dell'uso di modelli non adeguatamente testati, si producano risultati discriminatori, arbitrari o erronei, la cui ingiustizia o inesattezza non sarebbe, per di più, facilmente dimostrabile, né contestabile (stante l'opacità – *i.e.*,

¹¹ Si può affermare, in sintesi, che nella valutazione del merito di credito si è verificata una graduale estensione delle fonti di dati utilizzate (fenomeno che ha invero riguardato sia i modelli econometrici che le tecniche di *machine learning*), che oggi spaziano da dati finanziari strutturati (indicatori patrimoniali ed economico-finanziari, sull'andamento dei conti e dei pagamenti, di mercato) a dati non finanziari strutturati (dati di tipo socio-demografico ottenuti anche da fonti terze), dati non strutturati finanziari (analisi delle informazioni transazionali e di quelle derivate da *open banking*) e dati non strutturati non finanziari (dati di navigazione, *digital footprint*, informazioni conferite sui *social network*): AA. VV., *Intelligenza artificiale nel credit scoring. Analisi di alcune esperienze nel sistema finanziario italiano*, in BANCA D'ITALIA, *Questioni di economia e finanza*, 2022, n. 721, p. 15.

¹² La letteratura sul tema è ormai molto ampia. *Ex multis*, v. A. DAVOLA, *Algoritmi decisionali e trasparenza bancaria*, Milano, 2020, p. 136 ss.; F. MATTASSOGLIO, *Innovazione tecnologica e valutazione del merito creditizio del consumatore*, Milano, 2018; EAD., *La valutazione "innovativa" del merito creditizio del consumatore e le sfide per il regolatore*, in *Diritto della banca e del mercato finanziario*, 2020, n. 2, p. 187 ss.; G. L. GRECO, *Credit scoring 5.0*, cit., p. 74 ss.; L. AMMANNATI, G. L. GRECO, *Il credit scoring "intelligente": esperienze, rischi e nuove regole*, in *Rivista di diritto bancario*, 2023, n. 3, p. 461 ss.; M. RABITTI, *Credit scoring via machine learning e prestito responsabile*, *ivi*, 2023, n. 1, p. 175 ss.; P. GAGGERO, C. A. VALENZA, *Le moderne tecniche di credit scoring tra GDPR, disciplina di settore e AI Act*, *ivi*, 2024, n. 3, p. 825 ss.; A. M. G. GALATO, *Credit scoring e trasparenza algoritmica*, *ivi*, 2025, n. 1, p. 15; M. PROTO, *Contrattazione tecnobancaria e gestione dei dati: il consenso dell'interessato nella valutazione del merito creditizio*, in *Foro italiano*, 2025, n. 2, c. 114 ss.; N. M. FARAONE, *Spunti ricostruttivi in materia di profilazione e valutazione del merito creditizio nella nuova strategia europea dei dati*, in *Analisi Giuridica dell'Economia*, 2025, n. 1, p. 267; N. AGGARWAL, *The Norms of Algorithmic Credit Scoring*, in *Cambridge Law Journal*, 2021, 80, 1, 42 ss. (<https://ssrn.com/abstract=3569083>); N. REMOLINA, *The Role of Financial Regulators in the Governance of Algorithmic Credit Scoring*, SMU Centre for AI & Data Governance Research Paper, 2022, 2 (<https://ssrn.com/abstract=4057986>); D.K. CITRON, F. PASQUALE, *The Scored Society: Due Process for Automated Predictions*, in *Washington Law Review*, 2014, vol. 89 (1), p. 5 ss.

¹³ Per tali aspetti, si rimanda alle opere citate nella nota precedente.

¹⁴ Sul tema v. G. MATTARELLA, *L'inclusione finanziaria degli immigrati. La tutela del consumatore vulnerabile nei servizi bancari*, Torino, 2021, p. 55 ss.

l'inintelligibilità per la mente umana – dei processi decisionali adottati dai più complessi *predictive algorithms* e dei risultati cui gli stessi danno luogo¹⁵)¹⁶.

A favore delle nuove tecniche, inoltre, viene spesso rilevato come la più approfondita conoscenza del soggetto scrutinato, determinata dalla maggiore disponibilità di informazioni sul suo conto e dalla capacità della macchina di mettere in correlazione grandi quantità di dati disomogenei, non soltanto di natura finanziaria, permette valutazioni particolarmente accurate, che consentono di offrire al singolo individuo prodotti ritagliati sulle sue effettive esigenze, di diminuire i tempi di erogazione del credito e di applicare tassi di interesse più contenuti, riducendo, al contempo, il rischio di inadempimento degli obblighi restitutori.

In senso opposto, tuttavia, si obietta che l'accurata profilazione del cliente può sfociare in condotte aggressive o opportunistiche da parte degli intermediari, i quali, ad esempio, potrebbero sfruttare la propensione dello stesso a sostenere un costo più elevato per l'erogazione del prestito (applicando prezzi personalizzati), o approfittare delle condizioni di vulnerabilità economica o culturale di determinati soggetti, per indurli ad acquistare specifici prodotti, in ipotesi meno vantaggiosi di altri.

Per finire, non può tralasciarsi di osservare come l'uso massiccio di dati personali, funzionale all'accuratezza della prognosi di solvibilità del cliente, non soltanto pone evidenti problemi di tutela della *privacy*, ma incide negativamente anche sotto il delicato profilo dell'autodeterminazione degli individui, le cui scelte comportamentali potrebbero essere condizionate (e quindi limitate) dalla consapevolezza che qualunque informazione sul proprio conto, comprese quelle estraibili dalle innumerevoli tracce digitali quotidianamente disseminate in rete, può assumere rilevanza ai fini di una futura determinazione del merito creditizio, se è vero che, come è stato efficacemente affermato, «*all data is credit data*»¹⁷. Circostanza che si riflette, per converso, sull'eventualità che le procedure di concessione del credito siano soggette a pratiche di manipolazione da parte di soggetti con elevate abilità digitali, che – ove a conoscenza della logica di funzionamento dell'algoritmo predittivo – potrebbero modulare la propria condotta in modo da influenzare a proprio favore gli esiti del sistema di *credit scoring*, così pregiudicandone l'effettiva utilità¹⁸.

Il quadro, insomma, presenta molte zone d'ombra, che, considerata la già rilevata mancanza di dati di lungo periodo, impediscono di esprimersi con sicurezza sulla reale efficacia delle nuove metodologie di *scoring* algoritmico. Un fatto è certo, nondimeno: il crescente grado di diffusione delle medesime solleva interrogativi giuridici che non possono restare privi di risposta.

¹⁵ Per tutti, F. PASQUALE, *The Black Box Society. The Secret Algorithms That Control Money and Information*, Cambridge, Massachusetts, 2015.

¹⁶ In generale, sui c.d. pregiudizi algoritmici (di tipo cognitivo, legati a difetti concernenti la fase di raccolta dei dati o a discriminazioni implicite nel raggruppamento degli stessi, o di tipo statistico, imputabili a dati estrapolati da contesti affetti da discriminazione strutturale, come, ad es., in caso di uso di un campione parziale, tale da svantaggiare soggetti sottorappresentati), v. G. MALGIERI, G. COMANDÉ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, in *International Data Privacy Law*, vol. 7, 4, 2017, p. 248 ss.; D. IMBRUGLIA, *Le presunzioni delle macchine e il consenso dell'interessato*, in *Rivista trimestrale di diritto e procedura civile*, 2023, n. 3, p. 932 ss.; B. FRAGASSO, *Intelligenza artificiale e responsabilità penale. Principi e categorie alla prova di una tecnologia "imprevedibile"*, Torino, 2025, p. 32 ss.

¹⁷ La paternità della massima è attribuita a Douglas Merrill, fondatore della *tech company* statunitense ZestFinance.

¹⁸ G. COMANDÉ, *Leggibilità algoritmica e consenso al trattamento dei dati personali*, in *Danno e responsabilità*, 2022, n. 1, p. 41; AA. VV., *Intelligenza artificiale nel credit scoring*, cit., p. 19.

3. *Il credit scoring algoritmico come processo decisionale automatizzato (art. 22 GDPR). Il caso Schufa (CGUE, 7 dicembre 2023, C-634/21)*

A tali esigenze ha provveduto, in parte, la decisione resa dalla Corte di Giustizia il 7 dicembre 2023 nella causa C-634/21 (*Schufa*)¹⁹.

La pronuncia ha sollevato immediato interesse nella comunità scientifica, poiché con essa la Corte di Giustizia ha affrontato per la prima volta lo spinoso tema dei processi decisionali interamente automatizzati concernenti le persone fisiche, disciplinati dal Regolamento UE 2016/679 sulla protezione dei dati (GDPR), all'art. 22 («Processo decisionale automatizzato relativo alle persone fisiche, compresa la profilazione»). Invero, la progressiva emersione di sistemi decisionali funzionanti senza necessità di intervento umano, connessa allo sviluppo tecnologico degli ultimi anni, sembrava inevitabilmente destinata a sfociare in un contenzioso, sorto proprio in relazione ad un servizio di valutazione automatizzata del merito creditizio dei consumatori, fornito da una società di diritto tedesco (*Schufa*) ai propri *partner* contrattuali (perlopiù banche e intermediari finanziari, che impiegano tali informazioni a supporto delle decisioni di concessione del credito).

Più in dettaglio, *Schufa* è un ente in grado di formulare dei pronostici automatizzati in merito alle probabilità di rimborso di un prestito da parte di una determinata persona, in virtù dell'assegnazione della medesima ad una classe di individui aventi caratteristiche simili (operazione, tecnicamente consistente in una profilazione²⁰, che permette di prevedere il comportamento futuro dell'interessato sulla base di calcoli di tipo statistico)²¹. Tale procedura acquista dunque rilevanza nell'ottica della protezione dei dati personali del soggetto valutato, poiché il calcolo automatizzato del punteggio (*scoring*) di solvibilità del richiedente il prestito postula, appunto, l'analisi di dati e informazioni personali, necessari per la profilazione; dati che l'interessato, a seguito del diniego di un finanziamento da parte di un ente creditizio, basato su informazioni negative trasmesse all'intermediario da *Schufa*, pretendeva gli fossero da quest'ultima comunicati, anche al fine di poter ottenere la rettifica di eventuali errori.

¹⁹ CGUE, 7 dicembre 2023, causa C-634/21, *Schufa Holding (Scoring)*, pubblicata in *Banca Borsa Titoli di Credito*, 2024, n. 6, II, p. 715, con nota di A.G. GRASSO, *Decisioni automatizzate e merito creditizio: la Corte di Giustizia sul credit scoring*, e in *Nuova giurisprudenza civile commentata*, 2024, n. 2, p. 416, con nota di F. D'ORAZIO, *Il credit scoring e l'art. 22 del GDPR al vaglio della Corte di Giustizia*. Per completezza, si rammenta che, nella stessa data, la Corte ha pronunciato anche la sentenza relativa alle cause C-26/22 e C-64/22, *Schufa Holding (Esdebitazione)*.

²⁰ Il concetto di profilazione è definito dall'art. 4, par. 4, GDPR come «qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica». La profilazione si caratterizza, dunque, per il peculiare obiettivo di valutare aspetti personali di un soggetto sulla base di un trattamento automatizzato di dati personali (sicché non v'è profilazione in presenza di una semplice classificazione di individui, ad es. per sesso o per età, ove priva di una valutazione delle loro caratteristiche personali volta a trarre particolari conclusioni).

²¹ In pratica, il calcolo dei punteggi di solvibilità (*scoring*) si basa sul presupposto che sia possibile prevedere il comportamento futuro di una persona (nella specie, l'adempimento dei propri debiti), assegnandola ad una classe di individui con caratteristiche analoghe che si siano già comportati in un certo modo. Per effettuare le proprie previsioni, dunque, il sistema combina dati del passato della persona valutata (ad es., informazioni reddituali, dati tratti dai conti bancari, dati relativi alla storia creditizia dell'interessato, ecc.) con l'inserimento della stessa all'interno di una categoria di soggetti aventi analogo profilo, presumendosi che, in situazioni standardizzate, gli individui tendano a comportarsi nello stesso modo.

Schufa, tuttavia, indicava al privato solamente il punteggio attribuitogli e, in modo generico, le relative modalità di calcolo, rifiutandosi di esplicitare le informazioni poste a base dello *scoring* e la relativa ponderazione, in quanto coperte da segreto commerciale. Obiettava, inoltre, che la decisione finale di non erogare il credito era stata comunque adottata dalla propria controparte contrattuale (il finanziatore), cui *Schufa* si era limitata a fornire le informazioni commerciali funzionali alla decisione medesima.

Nella specie, pertanto, si poneva il problema di determinare se il calcolo del tasso di probabilità di rimborso dei prestiti operato da *Schufa* costituisse un processo decisionale totalmente automatizzato relativo alle persone fisiche, atteso che tale attività, oggetto di un generale divieto *ex art.* 22, par. 1, GDPR, è ammessa solo in presenza di specifiche circostanze²², e fatta comunque salva l'adozione di efficaci misure di salvaguardia dei diritti, delle libertà e degli interessi legittimi del soggetto sottoposto al trattamento²³.

Ne discende che, ove inquadrata nell'art. 22 GDPR, l'attività in esame risulterebbe legittima solo in presenza di una delle situazioni derogatorie previste dall'art. 22, par. 2, cit.²⁴, fermo restando, per altro verso, che in tal caso *Schufa* sarebbe tenuta a dare accesso alle informazioni supplementari circa l'esistenza di un processo decisionale automatizzato, sulla logica utilizzata, nonché sull'importanza e le conseguenze del predetto trattamento, cui l'interessato ha diritto ai sensi dell'art. 15, par. 1, lett. *b*), GDPR.

Viceversa, ove dovesse ritenersi che, in situazioni come quella in esame, soltanto al provvedimento assunto dal terzo finanziatore potrebbe attribuirsi, se del caso, la qualifica di decisione adottata in base ad un trattamento automatizzato relativo alle persone fisiche, *Schufa* non dovrebbe garantire alcun accesso alle informazioni di cui all'art. 15 cit., non essendo alla stessa imputabile l'adozione di alcun atto rilevante ai sensi dell'art. 22 cit.

Da qui, dunque, la questione pregiudiziale sollevata innanzi alla Corte, con la quale il giudice del rinvio ha chiesto di accertare se l'art. 22, par. 1, GDPR debba essere interpretato nel senso che costituisce un «processo decisionale automatizzato relativo alle persone fisiche», ai sensi di tale disposizione, il calcolo automatizzato, da parte di una società che fornisce informazioni commerciali, di un tasso di probabilità basato su dati personali relativi a una persona e riguardante la capacità di quest'ultima di onorare impegni di pagamento in futuro, qualora da tale tasso di probabilità dipenda in modo decisivo la stipula, l'esecuzione

²² In sintesi, l'art. 22 GDPR prevede che l'interessato «ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona» (par. 1), a meno che tale decisione «a) sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento; b) sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato; c) si basi sul consenso esplicito dell'interessato» (par. 2). Sulla portata delle eccezioni al divieto generale di processi decisionali automatizzati, con specifico riferimento alle attività di *credit scoring* algoritmico, v. P. GAGGERO, C. A. VALENZA, *Le moderne tecniche di credit scoring*, cit., p. 849 ss.

²³ Ci si riferisce, in particolare, alla necessità di garanzie atte ad assicurare un trattamento equo e trasparente dei dati dell'interessato, comprensive del diritto ad ottenere l'intervento umano, ad esprimere la propria opinione e a ottenere una spiegazione della decisione conseguente alla valutazione, nonché all'impiego di procedure matematico-statistiche appropriate per la profilazione e di misure organizzative volte a ridurre al minimo il rischio di errori, a garantire la sicurezza dei dati personali e ad impedire eventuali discriminazioni (GDPR, considerando n. 71).

²⁴ Nel caso di specie rileverebbe, quale fonte autorizzativa della decisione robotizzata, il diritto dello Stato membro di residenza dell'interessato (art. 22, par. 2, lett. *b*)), ed in particolare l'art. 31 del *Bundesdatenschutzgesetz* (legge federale sulla protezione dei dati del 30 giugno 2017, BDSG), intitolato «Protezione delle operazioni economiche in caso di "scoring" e di informazioni sulla solvibilità».

o la cessazione di un rapporto contrattuale con tale persona da parte di un terzo, al quale è comunicato tale tasso di probabilità.

Ebbene, nell'affrontare il quesito, la Corte ha avallato – correttamente, ad avviso di chi scrive – un'interpretazione estensiva dell'art. 22 GDPR, che appare vieppiù conforme alla *ratio* della norma e che corrobora ulteriormente la protezione rafforzata dalla stessa offerta ai diritti dell'interessato. È appena il caso di precisare, invero, che – stabilendo rigorosi requisiti di liceità, accompagnati da specifici obblighi di informazione a carico del titolare del trattamento – le disposizioni in esame mirano a tutelare le persone fisiche dai rischi insiti nei processi decisionali automatizzati, ritenuti intrinsecamente pericolosi per l'interessato, anche in ragione degli effetti discriminatori che dagli stessi potrebbero scaturire²⁵. Rischio non disgiunto, peraltro, dall'ulteriore pericolo di compressione delle possibilità di scelta degli individui, per effetto di processi di segregazione degli stessi all'interno di specifiche categorie; di isolamento del soggetto dalle informazioni che non risultano conformi al profilo che dello stesso viene tracciato; di dissuasione dei singoli dal tenere condotte che potrebbero apparire censurabili o controproducenti, ecc., con effetti distorsivi che incidono in modo tutt'altro che indifferente sulla vita delle persone e, in termini più generali, sul funzionamento stesso delle società democratiche²⁶.

Proprio l'esigenza di assicurare una tutela più ampia possibile ai diritti delle persone fisiche, dunque, legittima l'adozione di un criterio interpretativo dell'art. 22 cit. orientato teleologicamente a garantire il perseguimento delle finalità proprie di tale disposizione, nonché, più in generale, del provvedimento normativo in cui la stessa si trova collocata, stante il principio, richiamato nella stessa decisione in commento, secondo cui l'interpretazione di una disposizione del diritto dell'Unione richiede che si tenga conto non soltanto del suo tenore letterale, bensì anche «del contesto in cui essa si inserisce nonché degli obiettivi e della finalità che persegue l'atto di cui essa fa parte» (punto 41).

Seguendo questa logica, la Corte, dopo avere premesso che l'applicabilità dell'art. 22 cit. presuppone l'esistenza di tre condizioni cumulative (i.e., l'esistenza di una «decisione», il fatto che la stessa sia «basata unicamente sul trattamento automatizzato, compresa la profilazione», infine che la decisione debba produrre «effetti giuridici» sull'interessato, o incidere «in modo analogo significativamente sulla sua persona»), si sofferma sul significato di ciascuna di esse, precisando innanzitutto che il termine «decisione» può ritenersi comprensivo, in mancanza di una puntuale definizione normativa, di qualsiasi misura che risulti atta ad incidere, nei più svariati modi, sull'interessato. Come rilevato in dottrina, dunque, l'espressione va intesa in senso ampio e atecnico, identificando, indipendentemente

²⁵ In dottrina sono stati evidenziati i rischi latenti di discriminazioni insiti nei sistemi di *scoring* algoritmico, dovuti, ad es., al fatto che il programmatore potrebbe omettere di ponderare adeguatamente le diverse variabili rilevanti ai fini della definizione della specifica formula di meritevolezza creditizia da applicare alla clientela, o all'eccessiva ampiezza (in termini di quantità e varietà) dei dati personali analizzabili, con conseguente incremento del rischio di «correlazioni spurie» (i.e., accostamenti arbitrari tra serie di dati privi di significato in termini di reale interdipendenza delle variabili considerate) da parte degli algoritmi di *machine learning* e di risultati irrazionali o poco trasparenti, dei quali l'interessato risulta peraltro impossibilitato a valutare la correttezza, l'equità e la legittimità (C. TABARRINI, *Comprendere la "Big Mind": il GDPR sana il divario di intelligibilità uomo-macchina*, in *Diritto dell'informazione e dell'informatica*, 2019, n. 2, p. 560 ss.; A. OTTOLIA, *Big Data e Innovazione Computazionale*, in *Quaderni di Aida*, Torino, 2017, n. 28, p. 6 ss.; AA. VV., *Intelligenza artificiale nel credit scoring*, cit., p. 16 ss.).

²⁶ Su tali aspetti si v. E. PELLECCIA, *Profilazioni e decisioni automatizzate ai tempi della Black Box Society: qualità dei dati e leggibilità dell'algoritmo nella cornice della Responsible Research and Innovation*, in *Le nuove leggi civili commentate*, 2018, n. 5, p. 1212 ss.

dalla denominazione formale del provvedimento, tutte quelle determinazioni adottate per definire un procedimento (esclusi, dunque, gli atti preliminari, intermedi o interlocutori), che abbiano ripercussioni sulla persona interessata²⁷: ciò che si verifica, invero, anche nel caso del «calcolo della solvibilità di una persona sotto forma di tasso di probabilità relativo alla capacità di tale persona di onorare impegni di pagamento in futuro» (punto 46). Del resto, lo stesso considerando 71 del GDPR, nell'esemplificare la (ampia) portata dell'espressione «decisione», fa esplicito riferimento proprio al rifiuto automatico di una domanda di credito *on line*, richiamando una fattispecie non distante da quella in esame.

Nulla quaestio, invece, con riguardo alla seconda condizione sopra indicata (sussistenza di un trattamento interamente automatizzato, compresa la profilazione), essendo pacifico, nella prospettazione delle parti in causa, che l'attività di *Schufa* corrispondesse alla nozione di profilazione, per come definita dall'art. 4, punto 4, GDPR. È indubbio, d'altronde, che il calcolo automatizzato del tasso di solvibilità fosse basato, nella specie, sulla valutazione di dati personali, al fine di stimare la capacità di onorare l'eventuale debito contratto (ossia, un aspetto personale di una persona fisica, quale la sua affidabilità creditizia)²⁸.

Ben più interessanti, invece, le osservazioni sviluppate dalla Corte con riferimento alla terza e ultima condizione di applicabilità della norma, rappresentata dalla produzione, da parte della decisione automatizzata, di effetti giuridici sulla persona interessata, o comunque dalla significativa incidenza di siffatta decisione sulla persona medesima²⁹. Riconoscendo, infatti, l'importanza dei pronostici formulati da *Schufa* nel contesto delle procedure di erogazione dei crediti (atteso che, come accertato anche dal giudice del rinvio, uno *scoring* negativo rappresenta condizione in linea di massima sufficiente per rifiutare il finanziamento), la Corte giunge ad affermare che siffatte decisioni debbano essere considerate rilevanti ai fini

²⁷ In questi termini, A.G. GRASSO, *Decisioni automatizzate e merito creditizio*, cit., p. 732.

²⁸ Ciò che merita aggiungere, semmai, è che, per comune opinione, l'avverbio «unicamente» (riferito al trattamento automatizzato dei dati, quale base esclusiva della decisione) non preclude l'applicazione dell'art. 22 GDPR alle decisioni che presuppongono forme marginali o meramente simboliche di coinvolgimento umano, prive, cioè, di un'effettiva influenza sull'esito del processo decisionale automatizzato. L'intervento umano che sterilizza l'applicazione della norma, infatti, è quello dotato di un'effettiva incidenza sulla decisione, effettuato da un soggetto che, esercitando la propria capacità di giudizio critico, abbia l'autorità e la competenza per modificarla (Gruppo di Lavoro Articolo 19, *Linee guida sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione ai fini del regolamento 2016/679*, 3 ottobre 2017, mod. 6 febbraio 2018). In dottrina, pertanto, era stato già osservato come nella fattispecie in esame potesse rientrare anche l'ipotesi del recepimento passivo di uno *scoring* creditizio algoritmico da parte di un soggetto umano che si limiti ad abbinare ad un determinato punteggio particolari condizioni contrattuali (E. PELLECCIA, *op. cit.*, pp. 1225-6). Caso analogo, quest'ultimo, a quello esaminato nella decisione *Schufa*, posto che, dagli accertamenti di fatto eseguiti dal giudice del rinvio, era emerso che l'attribuzione al consumatore di un punteggio insufficiente comportava, di fatto, il rifiuto della banca di concedere il prestito richiesto (punto 48).

²⁹ Là dove, per decisione idonea a produrre effetti giuridici si intende quella che incide sui «diritti giuridici di una persona, quali la libertà di associarsi ad altre persone, di votare nel contesto di un'elezione o di intraprendere azioni legali» o che influisce «sullo status giuridico di una persona o sui suoi diritti ai sensi di un contratto», mentre, per decisione idonea a incidere in modo analogamente significativo sulla persona si intende quella decisione in grado di «incidere in maniera significativa sulle circostanze, sul comportamento o sulle scelte dell'interessato; avere un impatto prolungato o permanente sull'interessato; o nel caso più estremo, portare all'esclusione o alla discriminazione di persone» (*Linee guida sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione ai fini del regolamento 2016/679*, cit.). A quest'ultimo proposito, vengono menzionate in via esemplificativa le decisioni che incidono sulla situazione finanziaria delle persone (influenzando, ad es., l'accesso al credito), o sull'accesso ai servizi sanitari, all'istruzione o all'impiego (sebbene queste ultime riguardino, a ben vedere, situazioni giuridiche soggettive già qualificabili come veri e propri diritti, ai sensi del diritto dell'Unione), ossia decisioni che, per le sensibili ripercussioni sull'interessato, assumono un'importanza paragonabile a quella delle decisioni giuridiche *stricto sensu*.

dell'art. 22 cit. anche se trasmesse alla banca che abbia poi negato il prestito, le stesse svolgendo un ruolo decisivo nella determinazione finale assunta dall'ente creditizio. La Corte adotta, in breve, un approccio sostanzialistico, valorizzando il passaggio procedurale (la formulazione automatizzata dello *scoring*) che, per il peso in concreto rivestito, determina l'adozione del provvedimento di diniego del credito, così incidendo sulla persona sottoposta al trattamento. Il punto appare, dunque, estremamente significativo e gravido di conseguenze pratiche, per la rilevanza che assume con riferimento ai casi, assai frequenti nella pratica, di "segmentazione" del trattamento automatizzato di dati fra più soggetti³⁰.

Può dirsi, pertanto, che la pronuncia riflette, sotto tale aspetto, le inquietudini dottrinarie circa la «travolgente forza pratica dell'algoritmo», espressa dalla capacità dei sistemi di valutazione automatizzata, ove introdotti all'interno di un processo decisionale umano, di "catturare" - per mere ragioni di convenienza pratica, legate, ad es., all'alleggerimento dell'onere di motivare i provvedimenti e alla (presunta) qualifica di scientificità e di oggettività attribuibile all'*output* del sistema - la decisione dell'agente umano, neutralizzando il principio normativo di "non esclusività algoritmica"³¹.

D'altra parte, pare difficile non concordare con le conclusioni della Corte, ove si osservi che - come puntualizzato dagli stessi giudici di Lussemburgo - una diversa e più restrittiva opzione ermeneutica, che qualificasse come «decisione» ex art. 22 cit. solo l'atto di rifiuto del credito adottato dal terzo (e non anche il precedente calcolo del tasso di solvibilità), comporterebbe un elevato rischio di elusione della norma e, conseguentemente, un vuoto di tutela del consumatore. Per un verso, infatti, accedendo a tale soluzione uno *scoring* algoritmico come quello effettuato nel caso in esame esulerebbe dal raggio di applicazione dell'art. 22, malgrado si tratti di provvedimento interamente automatizzato, idoneo ad incidere in modo significativo sull'interessato (nella misura in cui, come si è detto, tale provvedimento condiziona in modo preponderante le scelte altrui di erogazione del credito). Per l'altro, il soggetto interessato non riuscirebbe in alcun modo ad esercitare il suo diritto di accesso alle informazioni di cui all'art. 15, par. 1, GDPR (*i.e.*, informazioni sulla logica utilizzata e sulle conseguenze per l'interessato), posto che l'ente che fornisce il calcolo del tasso di solvibilità, pur disponendo delle predette informazioni, non sarebbe tenuto a soddisfare la relativa richiesta (non essendo considerato l'autore di una misura in grado di incidere sull'interessato), mentre l'ente finanziatore (cui verrebbe invece imputata, formalmente, la decisione basata sul trattamento automatizzato dei dati) non sarebbe in grado di ottemperarvi, non disponendo, di norma, delle necessarie informazioni (punti 60-62).

³⁰ F. D'ORAZIO, *Il credit scoring e l'art. 22 del GDPR*, cit., p. 421. La pronuncia dovrebbe dunque determinare l'assoggettamento all'art. 22 GDPR di tutte quelle società che, senza erogare direttamente credito, offrono a terzi servizi automatizzati di *scoring*. Si pensi, ad es., alle società di informazione creditizia (SIC) italiane, con riferimento alle quali, peraltro, l'apposito *Codice di condotta* adottato dal Garante della privacy con deliberazione del 6 ottobre 2022, afferma che le stesse, anche ove trattino dati personali mediante processi decisionali interamente automatizzati, non adottano "decisioni" in grado di incidere sui diritti e sulle libertà degli interessati (F. MATTASSOGLIO, *La Corte di giustizia europea, algoritmi e credit scoring. L'apertura del vaso di Pandora delle società che si "limitano" a elaborare gli scoring*, in *Dialoghi di diritto dell'economia*, gennaio 2025, p. 10).

³¹ A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal*, 2019, n. 1, p. 81. Discorre di una possibile forma di deferenza dell'uomo rispetto ai risultati offerti dall'IA, S. FOÀ, *Intelligenza artificiale e cultura della trasparenza amministrativa. Dalle "scatole nere" alla "casa di vetro?"*, in *Diritto amministrativo*, 2023, n. 3, p. 542, indicando tra gli effetti negativi del fenomeno (noto come *automation bias*, *anchoring effect* o anche come *effet mountainier*) la deresponsabilizzazione del soggetto chiamato ad assumere la decisione.

Un ultimo aspetto, per completezza, pare degno di nota. Dopo avere appurato che lo *scoring* operato da *Schufa* integra un processo decisionale automatizzato ex art. 22 GDPR, la Corte afferma che spetta al giudice del rinvio verificare l'esistenza di una base giuridica idonea ad autorizzare, ai sensi dell'art. 22, par. 2, lett. b), GDPR, l'adozione di una simile decisione, evidenziando altresì che, se la norma nazionale dovesse risultare, in esito a tale disamina, incompatibile con il diritto dell'Unione³², *Schufa* non soltanto avrebbe agito senza alcun fondamento giuridico, ma avrebbe violato *ipso iure* anche il divieto di cui all'art. 22, par. 1, GDPR. Tale accertamento, tuttavia, non risulterà di per sé sufficiente. Al giudice del rinvio, invero, competerà verificare anche il rispetto delle ulteriori condizioni previste dall'art. 22, par. 2, lett. b) (ossia l'esistenza di adeguate misure a tutela dei diritti, delle libertà e di legittimi interessi dell'interessato, compreso il diritto di ottenere l'intervento umano, di esprimere la propria opinione e di contestare la decisione) e par. 4 (ai cui sensi le decisioni di cui al par. 2 non possono basarsi, di norma, sui dati "sensibili" di cui all'art. 9), nonché degli artt. 5 e 6 GDPR, che stabiliscono, com'è noto, i principi di fondo (liceità, trasparenza, specificazione delle finalità della raccolta, minimizzazione, ecc.) cui è subordinato qualsiasi trattamento di dati. In nessun caso, infatti, gli Stati membri sarebbero legittimati a adottare, ai sensi dell'art. 22, par. 2, lett. b), norme che autorizzino la profilazione in spregio dei requisiti stabiliti dai sopra citati artt. 5 e 6, per come interpretati dalla giurisprudenza della Corte (punti 67 e 68).

4. (*Segue*). *La tutela della persona nel processo decisionale automatizzato. Il diritto ad ottenere «significative informazioni» sulla logica della decisione algoritmica*

La tutela della persona umana, in caso di processo decisionale automatizzato (reso lecito dalla presenza di una delle circostanze in deroga al divieto generale), appare dunque affidata dal GDPR alla previsione di garanzie e cautele di varia natura, che includono, oltre al diritto di ottenere l'intervento umano (in una logica avversa ad ogni forma di subalternità dell'uomo al potere decisionale della macchina), specifici obblighi volti ad assicurare la piena correttezza e trasparenza del trattamento dei dati e delle sue finalità.

In quest'ottica, il legislatore europeo ha previsto obblighi di *disclosure* che accompagnano, in ogni sua fase, l'intero svolgimento del processo decisionale, stabilendo, oltre all'obbligo del titolare del trattamento di comunicare all'interessato una dettagliata serie di informazioni, in concomitanza con l'ottenimento dai dati utilizzati (artt. 13 e 14 GDPR), il diritto di accesso dell'interessato ai dati e alle informazioni che lo riguardano, comprese, come già accennato, «informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento» (art. 15 GDPR).

Una certa dose di ambiguità, presente nel testo normativo, ha tuttavia finito per ingenerare dubbi in ordine alla portata dei diritti riconosciuti al soggetto sul quale incide la

³² Secondo il giudice del rinvio, invero, sussisterebbero seri dubbi circa la compatibilità dell'art. 31 BDSG con l'art. 22 GDPR, in quanto il legislatore tedesco disciplinerebbe solo «l'uso» di un tasso di probabilità come quello di cui trattasi nel procedimento principale (cioè, un tasso di probabilità di un certo comportamento futuro di una persona fisica allo scopo di decidere sulla stipulazione, sull'attuazione o sulla cessazione di un contratto con tale persona), e non il calcolo in quanto tale di tale tasso (punto 25). È tuttavia palese che la linea di demarcazione tra la mera determinazione delle specifiche condizioni in presenza delle quali è lecito fare ricorso allo *scoring* di solvibilità di una persona fisica e l'autorizzazione (implicita) all'adozione di tali decisioni automatizzate sia davvero evanescente.

decisione automatizzata, alimentati specialmente dalla mancanza di puntuali indicazioni relative alle caratteristiche (qualitative e quantitative) delle «informazioni significative» che l'interessato ha diritto di ottenere ai sensi dell'art. 15 cit.

Il nocciolo della questione – che la decisione *Schufa* per vero non affronta, in quanto esorbitante dai quesiti sottoposti alla Corte – attiene, in estrema sintesi, all'esistenza di un diritto dell'interessato alla spiegazione della decisione algoritmica, nonché alle modalità idonee a darvi, eventualmente, concreta attuazione.

Basandosi su argomenti di tipo storico-letterale, infatti, autorevole dottrina è giunta ad affermare che il GDPR riconoscerebbe all'interessato, al più, un diritto ad essere informato (solo) in merito alla logica generale e astratta di funzionamento del sistema automatizzato, ma non anche ad ottenere una vera e propria spiegazione della specifica decisione che lo riguarda³³. Come desumibile dall'identica formulazione dei rispettivi enunciati normativi, invero, l'art. 15 garantirebbe all'interessato l'accesso alle medesime informazioni contemplate dagli artt. 13 e 14, ossia quelle disponibili nella fase di avvio del processo decisionale, come tali non riferibili ad una misura non ancora adottata.

Tale ricostruzione, tuttavia, è stata convincentemente smentita da coloro che ritengono che le finalità protettive dell'art. 15 cit., quale norma che concorre ad assicurare l'obiettivo della piena trasparenza delle decisioni automatizzate, verrebbero vanificate ove l'interessato non fosse messo in condizione di comprendere, mediante informazioni espresse in un linguaggio chiaro e fruibile, l'intero percorso logico seguito dalla macchina e le motivazioni della singola decisione adottata (*i.e.*, *come* e *perché* il sistema sia giunto a quello specifico risultato)³⁴. Se così non accadesse, infatti, all'interessato verrebbe precluso l'effettivo esercizio dei diritti riconosciutigli dall'art. 22 GDPR, e segnatamente il diritto di esprimere la propria opinione e di contestare la decisione, in quanto logicamente fondati sulla piena comprensione delle ragioni alla stessa sottese. In altri termini, la *ratio* di tutela della persona umana che accomuna gli artt. art. 13, 14 e 15 del GDPR, impone di fornire all'interessato, in un formato allo stesso intelligibile, informazioni che permettano sia la comprensione *ex ante* del meccanismo decisionale automatizzato (e, dunque, delle finalità e delle caratteristiche tecnico-operative del sistema, dei criteri utilizzati per giungere alle decisioni e della relativa ponderazione, delle tipologie di dati analizzabili, ecc.), sia la spiegazione *ex post* della decisione algoritmica (con specificazione di dati, criteri e ragioni che hanno portato ad assumere uno specifico provvedimento, affinché possano esserne verificate la correttezza e la legittimità ed esercitato, se del caso, il diritto di contestazione)³⁵.

La trasparenza diviene, in tal modo, lo strumento giuridico attraverso il quale la persona può «comprendere la comprensione che la macchina ha della persona medesima nell'ambito di processi decisionali diretti ad incidere sulla sua sfera giuridica e vitale», innescando un circuito informativo che consente all'uomo di verificare che la decisione

³³ L. FLORIDI, S. WATCHER, B. MITTELSTADT, *Why a right to Explanation on Automated Decision-Making Does Not Exist in the GDPR*, in *International Data Privacy Law*, 2017, p. 76 ss.; in arg. v. anche G. FINOCCHIARO, *Intelligenza artificiale. Quali regole?*, Bologna, 2024, p. 83, secondo la quale il “diritto alla spiegazione” riceverebbe, a livello normativo, un riconoscimento formulato in modo ancora embrionale.

³⁴ G. MALGIERI, G. COMANDÈ, *op. cit.*, p. 248 ss., che ravvisano un diritto alla “leggibilità” dell'algoritmo, comprensivo dei concetti sia di trasparenza - intesa come disponibilità di informazioni dettagliate ed esaustive - che di comprensibilità delle logiche con cui vengono inferite le conclusioni; conf., G. CARAPEZZA FIGLIA, *Decisioni algoritmiche tra diritto alla spiegazione e divieto di discriminare*, in *Persona e mercato*, 2023, n. 4, p. 643; P. GAGGERO, C. A. VALENZA, *Le moderne tecniche di credit scoring*, cit., p. 845 ss.

³⁵ Diffusamente, F. BRAVO, *Trasparenza del codice sorgente e decisioni automatizzate*, in *Diritto dell'informazione e dell'informatica*, 2020, n. 4-5, pp. 712-9; C. TABARRINI, *Comprendere la “Big Mind”*, cit., p. 562 ss.

automatizzata – ottenuta in esito a quello che può essere definito come un «processo di eterodeterminazione tecnologica della sua identità», operato per via algoritmica - non risulti in alcun modo lesiva dei suoi diritti e libertà fondamentali³⁶.

5. (*Segue*). *Diritto alla spiegazione della decisione automatizzata e full disclosure dell'algoritmo. Il caso D&B (CGUE, 27 dicembre 2025, C-203/22)*

Il riconoscimento di un diritto alla spiegazione della decisione algoritmica (cioè, ad un'informativa non limitata ai dati di *input* e ai criteri generali di funzionamento del sistema automatizzato, ma estesa anche all'*output* del sistema stesso) sarebbe insufficiente, in difetto di indicazioni circa ciò che deve essere concretamente rappresentato all'interessato per assicurarne l'effettività. Il rischio, infatti, è che gli obblighi di trasparenza imposti per via normativa si trasformino - come già ampiamente sperimentato in altri casi³⁷ - in inutili adempimenti burocratici, risultando del tutto inidonei, in concreto, a superare il problema dell'opacità che tendenzialmente connota le decisioni algoritmiche. Opacità che – merita ricordare - dipende non soltanto dalla naturale difficoltà di comprendere i sofisticati meccanismi operativi dell'algoritmo da parte di coloro che siano sprovvisti di adeguate competenze informatiche, ma anche da consistenti limiti di carattere tecnologico (essendo noto che, specie nel caso di ricorso a sistemi di *machine learning*, può risultare estremamente difficoltoso, persino per gli stessi programmatori, ricostruire il percorso logico che ha condotto la macchina ad adottare una certa decisione³⁸) e normativo (si pensi alle disposizioni in materia di proprietà intellettuale o di segreto industriale, che, assicurando protezione alla formula matematica dell'algoritmo, vengono spesso invocate dai titolari del trattamento – proprio come accaduto nel caso *Schufa* – per opporsi alle richieste di informazioni sul funzionamento dei *software*).

Tali osservazioni mettono in luce, dunque, gli ostacoli che si frappongono ad una piena ed effettiva attuazione del principio normativo di trasparenza, a detrimento della tutela della persona interessata dalla decisione automatizzata. Affinché tale principio non rimanga lettera morta, infatti, occorre agire contestualmente su più piani, cercando di conseguire i seguenti risultati: in primo luogo, la decisione deve essere tecnicamente spiegabile (ciò che postula il ricorso a modelli totalmente interpretabili, che consentano di superare il problema dell'opacità intrinseca dei sistemi di *machine learning*); in secondo luogo, le informazioni rese all'interessato devono essere espresse in un linguaggio intelligibile da parte del destinatario e

³⁶ Sul tema, si vedano le approfondite riflessioni di R. MESSINETTI, *La tutela della persona umana versus l'intelligenza artificiale. Potere decisionale dell'apparato tecnologico e diritto alla spiegazione della decisione automatizzata*, in *Contratto e impresa*, 2019, n. 3, p. 867 ss., da cui sono tratte le espressioni riportate nel testo.

³⁷ È scontato il riferimento al problema dell'eccesso di informazioni (obbligatorie *ex lege*) che accompagna, in genere, la prestazione dei servizi del ramo finanziario (c.d. *information overload*), essendo noto che tale mole di notizie, molto dettagliate e di natura tecnica, non garantisce l'effettiva comprensione, da parte degli utenti (specie non professionali), delle effettive caratteristiche e del grado di rischio dei prodotti sottoscritti.

³⁸ È l'arcinoto problema dell'imperscrutabilità dell'*output* fornito da sistemi algoritmici complessi, efficacemente espresso dalla metafora della "scatola nera" (F. PASQUALE, *The Black Box Society*, cit.). Invero, nel caso del *machine learning*, l'algoritmo non applica razionalmente, secondo un metodo deduttivo, le regole e i principi dettati dal programmatore, ma apprende in modo autonomo, sulla base dell'esperienza, le regole del proprio funzionamento, essendo in grado di stabilire correlazioni tra i dati di *input* e di ricavare delle generalizzazioni, costruendo un modello che gli consente di offrire la soluzione ad un determinato quesito o problema.

risultare utili alla comprensione del sistema decisionale algoritmico, prendendo corpo in tali requisiti l'attributo di "significatività" che la legge associa al corredo informativo *ex art.* 15 GDPR; infine, occorre individuare un giusto punto di equilibrio tra il diritto di accesso a dati e informazioni da parte dell'interessato e i diritti altrui (con particolare riferimento ai diritti di privativa sul *software*, che precludono la rivelazione dell'algoritmo), attuando, come indicato dal GDPR, un equo bilanciamento tra esigenze contrapposte³⁹.

Orbene, una prima risposta alle questioni sopra rappresentate giunge dalla recentissima pronuncia della Corte di Giustizia del 27 febbraio 2025 nella causa C-203/22 (*D&B*), con la quale, a integrazione delle statuizioni contenute nella decisione *Schufa*, sono stati forniti utili criteri di interpretazione dell'art. 15 GDPR, ancora una volta con riferimento ad un servizio di *scoring* algoritmico.

La vicenda può essere così riassunta. A seguito di una controversia scaturita dalle doglianze di una cittadina austriaca, cui era stata rifiutata la stipula di un contratto di telefonia mobile del modesto importo di 10 euro mensili, a causa di un giudizio di solvibilità negativo elaborato con mezzi automatizzati dalla società *D&B*, il giudice del rinvio chiedeva che la Corte indicasse i requisiti sostanziali che le informazioni di cui all'art. 15 cit. devono possedere per essere considerate sufficientemente significative e se, al tal fine, le stesse debbano comprendere anche le parti dell'algoritmo alla base della profilazione della persona interessata⁴⁰. La Corte veniva chiamata a stabilire, in sostanza, se quest'ultima abbia diritto ad una «spiegazione esaustiva della procedura e dei principi concretamente applicati per usare, con mezzi automatizzati, i dati personali relativi a tale persona al fine di ottenerne un determinato risultato, quale un profilo di solvibilità» (punto 38).

Ebbene, i giudici europei, tenuto conto sia delle diverse accezioni semantiche delle espressioni «informazioni significative» e «logica utilizzata»⁴¹, sia del contesto generale –

³⁹ Ai sensi del considerando n. 63 del GDPR, infatti, il diritto di accesso dell'interessato ai dati personali che lo riguardano «non dovrebbe ledere i diritti e le libertà altrui, compreso il segreto industriale e aziendale e la proprietà intellettuale, segnatamente i diritti d'autore che tutelano il software», fermo restando, tuttavia, che «tali considerazioni non dovrebbero condurre a un diniego a fornire all'interessato tutte le informazioni».

⁴⁰ Più in dettaglio: C.K. si vedeva negare, come già accennato, la sottoscrizione di un contratto di telefonia mobile di modesto valore, in ragione di una valutazione negativa della sua affidabilità creditizia, effettuata per via automatizzata dalla società *D&B*. L'autorità austriaca per la protezione dei dati personali ordinava a quest'ultima di comunicare a C.K. informazioni significative sulla logica utilizzata nel processo decisionale automatizzato, ma la *D&B* si opponeva in sede giudiziaria, invocando l'esistenza di un segreto commerciale, tutelato dalla direttiva *know-how* (art. 2, punto 1, dir. UE 2016/943 del Parlamento europeo e del Consiglio, dell'8 giugno 2016, sulla protezione del *know-how* riservato e delle informazioni commerciali riservate (segreti commerciali) contro l'acquisizione, l'utilizzo e la divulgazione illeciti, in *GUUE* L 157/1 del 15 giugno 2016). La Corte federale amministrativa austriaca riconosceva la violazione dell'art. 15, par. 1, lett. b), GDPR da parte della *D&B*, con decisione divenuta esecutiva ai sensi del diritto nazionale, ma la domanda di esecuzione forzata proposta da C.K. innanzi all'amministrazione comunale della città di Vienna veniva respinta, costringendo l'interessata a rivolgersi al Tribunale amministrativo di Vienna. Quest'ultimo, infine, decideva di sospendere il giudizio e di sottoporre alla Corte di Giustizia una serie di questioni pregiudiziali, articolate in più punti, attinenti alla corretta interpretazione dell'art. 15 cit. (in sintesi, come già rilevato, quali requisiti sostanziali devono sussistere affinché le informazioni previste da tale norma siano considerate significative e se, al tal fine, tali informazioni debbano estendersi anche alla formula dell'algoritmo di profilazione).

⁴¹ Si legge nelle motivazioni della pronuncia che le accezioni del concetto di «informazioni significative» divergono nelle varie versioni linguistiche del GDPR, privilegiando, in alcuni casi, l'aspetto della funzionalità o della pertinenza, in altri quello dell'importanza, mentre, in altri casi ancora, il termine utilizzato potrebbe intendersi riferito sia alla buona intelligibilità che ad una certa qualità di tali informazioni (punto 40). Il concetto andrebbe ricostruito, pertanto, ricorrendo alla complementarità dei diversi significati ricavabili dalle varie versioni linguistiche (punto 41).

improntato al principio di trasparenza - nel quale si trova collocato l'art. 15 cit., giungono alla conclusione che tale disposizione si riferisce a «qualsiasi informazione pertinente relativa alla procedura e ai principi di utilizzo dei dati personali al fine di ottenerne, con mezzi automatizzati, un determinato risultato, mentre l'obbligo di trasparenza richiede inoltre che tali informazioni siano fornite in forma concisa, trasparente, comprensibile e facilmente accessibile» (punto 50).

Privilegiando un approccio marcatamente funzionalistico, la Corte specifica quindi che il diritto di accesso di cui all'art. 15 GDPR deve essere definito, nella sua portata concreta, in conformità con lo scopo principale cui la norma è preordinata, che è quello di garantire l'esercizio, da parte dell'interessato, degli ulteriori diritti allo stesso riconosciuti per via normativa: il diritto alla rettifica, alla cancellazione e alla limitazione del trattamento (artt. 16, 17 e 18), all'opposizione al trattamento stesso (art. 21), ovvero, con riferimento al caso di un processo decisionale automatizzato, ad esprimere il proprio punto di vista e a contestare la decisione (art. 22) (punti 54 e 55). In questa logica, dunque, si rinviene espressamente nell'art. 15 cit. la base giuridica di un vero e proprio diritto alla spiegazione della decisione algoritmica (concernente, cioè, il «funzionamento del meccanismo alla base di un processo decisionale automatizzato di cui tale interessato è stato oggetto e sul risultato a cui detta decisione ha condotto»: punto 57), la cui concreta attuazione richiede la trasmissione di informazioni pertinenti, concise e, soprattutto, facilmente intelligibili da parte dell'interessato⁴².

La diretta conseguenza di tale approccio è l'affermazione secondo cui tali requisiti non possono ritenersi soddisfatti dalla mera comunicazione della formula matematica dell'algoritmo o da una dettagliata descrizione di tutte le fasi del procedimento decisionale automatizzato (trattandosi, invero, di informazioni sprovviste dei necessari caratteri di concisione e comprensibilità, specie se espresse in linguaggio informatico), dovendosi spiegare all'interessato, semmai, quali dei suoi dati personali siano stati utilizzati, ed in che modo, nell'ambito della specifica procedura robotizzata⁴³.

Quanto al riferimento alla «logica utilizzata» nel processo decisionale automatizzato, che costituisce l'oggetto di dette «informazioni significative», si rileva che l'espressione può coprire un ampio ventaglio di modalità di utilizzo di dati personali e di altri dati al fine di ottenere, con mezzi automatizzati, un determinato risultato. Ad esempio, nelle versioni in lingua ceca e polacca, vengono utilizzati dei termini che possono essere tradotti con «procedura» e «principi» (punto 42). Se ne ricava, in conclusione, che la formulazione dell'art. 15, par. 1, lett. b), del GDPR riguarda qualsiasi informazione pertinente relativa alla procedura e ai principi di utilizzo, con mezzi automatizzati, di dati personali, al fine di ottenerne un determinato risultato (punto 43).

⁴² La comunicazione di informazioni (chiare e concise, nonché complete e contestualizzate) relative alla profilazione effettuata dalla *DeB*, peraltro, consentirebbe all'interessata di verificarne anche la correttezza e la veridicità (punti 53 e 63), oltre che di ricostruire la correlazione (e dunque l'intrinseca coerenza) tra le informazioni trattate e la valutazione che ne è scaturita (v., in tal senso, le conclusioni dell'Avv. Generale De La Tour, presentate nella causa C-203/22 il 12 settembre 2024).

⁴³ Risultato che – precisa frettolosamente la Corte – si sarebbe potuto raggiungere, ad esempio, comunicando all'interessato «come una variazione a livello dei dati personali presi in considerazione avrebbe portato a un risultato diverso» (punto 62). La Corte sembrerebbe aderire, pertanto, all'idea secondo cui, ai fini del rispetto degli obblighi informativi sanciti dal GDPR, potrebbe essere sufficiente una c.d. spiegazione controfattuale, volta ad offrire ai soggetti interessati informazioni relative al più piccolo cambiamento che avrebbe dovuto essere apportato alla situazione reale per conseguire l'esito atteso o desiderato (ad es.: se Tizio avesse avuto un reddito di € 50.000, anziché di € 45.000, gli sarebbe stato concesso il credito che gli è stato invece negato). Si tratta, tuttavia, di un criterio piuttosto controverso, in quanto, come rilevato in dottrina, esso consente un'analisi meramente esterna del processo decisionale automatizzato (cioè, non basata sulla logica interna di funzionamento dell'algoritmo), inidonea a soddisfare concretamente i prescritti requisiti di significatività delle informazioni da trasmettere all'interessato (per la dimostrazione di tale assunto v. C. TABARRINI, *Comprendere la "Big Mind"*, cit., p. 577 ss.).

In altri termini, i dati rispetto ai quali può essere esercitato il diritto di accesso dell'interessato *ex art. 15 GDPR* (per ottenere, in sostanza, la spiegazione della decisione algoritmica) vengono individuati in quelli che rivestono concreta utilità rispetto all'esercizio degli altri diritti previsti dal medesimo provvedimento. La Corte, pertanto, conferma l'approccio in base al quale il flusso informativo in favore dell'interessato deve consentirgli di comprendere sia la logica generale seguita dal sistema che ha adottato la decisione, sia le specifiche motivazioni a fondamento di quest'ultima, includendovi, in sostanza, informazioni chiare e concise, ma al contempo esaustive, relative a modalità di funzionamento, criteri, parametri e variabili utilizzate dal sistema decisionale automatizzato, e non anche dati tecnico-informatici (l'algoritmo) difficilmente decifrabili. Il dato che non va comunicato, in conclusione, è solo quello che risulta inservibile – per complessità e inintelligibilità – ai fini di trasparenza e di tutela dei diritti individuali riconosciuti dal GDPR.

Alla luce di quanto osservato, possono facilmente intuirsi le ricadute di tali conclusioni nel contesto ordinamentale italiano, là dove la giurisprudenza, seguendo altri percorsi argomentativi, si è sovente mostrata incline ad affermare il principio della conoscibilità dell'algoritmo (affinché siano chiare, e quindi sindacabili, le regole in base alle quali è stato impostato il procedimento robotizzato e ottenuto il relativo *output*)⁴⁴, anche nell'ottica della valida prestazione da parte dell'interessato, a monte della decisione, del consenso al trattamento dei propri dati personali (quest'ultimo non potendosi ritenere realmente consapevole e informato, in caso di ignoranza dei criteri di funzionamento del processo decisionale automatizzato)⁴⁵.

È pur vero, nondimeno, che la questione relativa all'obbligo di divulgazione dell'algoritmo – escluso, in linea di principio, per le ragioni sopra illustrate – non può dirsi risolta dalla Corte in senso radicalmente negativo, riaffiorando comunque, nelle motivazioni

⁴⁴ La giurisprudenza amministrativa ha in più occasioni affermato che, in caso di ricorso a procedure algoritmiche da parte della P.A., deve essere garantita la conoscibilità dell'algoritmo (ossia di una regola espressa in un linguaggio diverso da quello giuridico) in tutti i suoi aspetti (autori, procedimento, meccanismo decisionale, dati selezionati come rilevanti, ecc.); al contempo, ha ribadito come la “caratterizzazione multidisciplinare” dell'algoritmo non esime dalla necessità che la relativa formula tecnica sia corredata da spiegazioni che la traducano nella regola giuridica ad essa sottesa e che la rendano comprensibile, senza che possa assumere rilievo, in senso contrario, la riservatezza dei meccanismi informatici, invocata dalle imprese produttrici (Cons. Stato, sez. VI, 8 settembre 2021, n. 6326; Cons. Stato, sez. VI, 4 febbraio 2020, n. 881; Cons. Stato, sez. VI, 13 dicembre 2019, n. 8472, in *Foro it.*, 2020, 6, III, 340 e Cons. Stato, sez. VI, 8 aprile 2019, n. 2270, *ibidem*, 2019, 11, III, 606). Sul punto si vedano, però, i rilievi critici di F. BRAVO, *op. cit.*, pp. 723-4, il quale ritiene che «l'ostensione del codice sorgente è misura di eccessivo pregiudizio del diritto fondamentale alla proprietà intellettuale, che invece può essere bilanciato in maniera proporzionale con il principio di trasparenza attraverso obblighi di informazione alternativi (...), maggiormente funzionali a rendere edotto l'interessato della “logica” utilizzata nella decisione algoritmica, accompagnati dall'individuazione di adeguate conseguenze giuridiche in caso di violazione, che lo tutelino, sul piano rimediabile, dagli effetti negativi della decisione automatizzata a lui pregiudizievole».

⁴⁵ Cass., 10 dicembre 2023, n. 28358, in *Il diritto di famiglia e delle persone*, 2023, n. 4, p. 1558, con nota di I. RAPISARDA, *Autodeterminazione informativa e trattamento algoritmico dei dati personali. Il caso del rating reputazionale*, e Cass., 24-25 marzo 2021, n. 14381, sulla quale v. G. COMANDÉ, *Leggibilità algoritmica*, cit., p. 33 ss. (entrambe riferite al noto caso *Mevaluate*, piattaforma *web* creata per calcolare, in modo asseritamente attendibile e imparziale, il *rating* reputazionale di persone fisiche e giuridiche). Merita precisare, tuttavia, che, a differenza della giurisprudenza amministrativa, la Cassazione non si spinge sino ad esigere l'ostensione della formula matematica dell'algoritmo, richiedendo invece, ai fini dell'effettiva libertà e consapevolezza del consenso al trattamento dei dati, che siano conoscibili lo «schema esecutivo dell'algoritmo» (*i.e.*, i passi da eseguire in sequenza, nel procedimento di risoluzione di un problema, per giungere ad un risultato) e gli elementi di cui lo stesso si compone, con ciò alimentando il problema della difficile traduzione in termini giuridici di concetti appartenenti alle scienze matematiche e informatiche.

della decisione in esame, il problema del bilanciamento fra diritto d'accesso dell'interessato e diritto d'autore sul *software*: i giudici di Lussemburgo, infatti, non escludono che, per poter ottemperare a quanto previsto dall'art. 15 cit., il titolare del trattamento si veda costretto a comunicare all'interessato alcuni dati di terzi protetti dal GDPR o dal segreto commerciale. Si noti, d'altronde, che, secondo il consulente tecnico nominato dal giudice del rinvio, il corretto adempimento degli obblighi informativi gravanti sulla *De'B* presupponeva che all'interessata fosse comunicata anche «la formula matematica alla base del calcolo che ha portato allo score controverso nel procedimento principale», unitamente ad un «elenco degli “scores” di persone («scoring») per i sei mesi precedenti e i sei mesi successivi all'attribuzione dello “score” di CK e che sono stati ottenuti sulla base della stessa regola di calcolo» (punti 24 e 25).

Persiste, quindi, l'esigenza di una composizione di interessi contrapposti. Ed è un problema che la Corte affronta muovendo dal presupposto che il diritto alla protezione dei dati personali non costituisce una prerogativa assoluta e va quindi temperato con altri diritti, in ossequio al principio di proporzionalità⁴⁶, per concludere che dovrà essere il giudice nazionale (o l'autorità amministrativa) a valutare quali dati personali delle parti o di terzi debbano essergli comunicati, in modo da poter ponderare, con cognizione di causa e nel rispetto del citato principio di proporzionalità, gli interessi in gioco. In altre parole, la portata del diritto di accesso dovrà essere stabilita, secondo le circostanze del caso, dall'autorità interpellata, che potrà decidere di autorizzare la divulgazione (integrale o parziale) all'interessato di certi dati (compreso l'algoritmo), nei limiti in cui ciò non comporti un eccessivo sacrificio dei diritti altrui (nella specie, la proprietà intellettuale o il segreto industriale).

La Corte adotta, in tal modo, una soluzione che può per certi versi apparire insoddisfacente (posto che, sposando la logica del caso per caso, lascia inevitabilmente spazio all'incertezza), ma che viene temperata, al contempo, dall'apertura verso il superamento di quello che è stato visto come un notevole limite alla tutela dei diritti dell'interessato: la segretezza dell'algoritmo preposto all'assunzione della decisione automatizzata⁴⁷. Dal tenore della pronuncia, infatti, si evince che l'algoritmo non deve ritenersi coperto da un velo assoluto di riservatezza, ma può anche essere divulgato, all'occorrenza, ove il giudice ritenga che ciò realizzi un equo bilanciamento degli interessi in conflitto. D'altra parte, come osservato in dottrina, essere ingiustamente considerati cattivi pagatori, per effetto dell'attribuzione di un punteggio di affidabilità creditizia negativo, integra una grave lesione della reputazione personale, sicché, al concretizzarsi di tale rischio, il diritto di accesso dell'interessato dovrebbe prevalere sulla tutela della proprietà industriale, in quanto posto a presidio del fondamentale principio, intimamente connesso alla reputazione, della dignità della persona⁴⁸.

⁴⁶ Esplicito, al riguardo, il considerando n. 4 del GDPR: «Il diritto alla protezione dei dati di carattere personale non è una prerogativa assoluta, ma va considerato alla luce della sua funzione sociale e va temperato con altri diritti fondamentali, in ossequio al principio di proporzionalità».

⁴⁷ V. ad es. F. MATTASSOGLIO, *La Corte di giustizia europea, algoritmi e credit scoring*, cit., p. 9, secondo la quale l'esclusione dell'algoritmo dagli obblighi informativi rischierebbe di depotenziare gli strumenti di tutela dell'interessato. Con riferimento agli algoritmi utilizzati dalla P.A., v. invece S. FOÀ, *Intelligenza artificiale e cultura della trasparenza amministrativa*, cit., p. 535.

⁴⁸ E. FALLETTI, *Alcune riflessioni sull'applicabilità dell'art. 22 GDPR in materia di scoring creditizio*, in *Diritto dell'informazione e dell'informatica*, 2024, n. 1, p. 115. Sul rapporto tra trattamento dei dati personali, reputazione e dignità della persona, v. G. CERRINA FERONI, *Intelligenza artificiale e sistemi di scoring sociale. Tra distopia e realtà*, *ivi*, 2023, n. 1, p. 12.

6. *Le ulteriori affermazioni normative del diritto alla spiegazione della decisione algoritmica. La direttiva UE 2023/2225 sul credito ai consumatori (CCD II) e il Regolamento UE 2024/1689 sull'intelligenza artificiale (AI Act).*

È indubbio che le recenti statuizioni della Corte di Giustizia, per come sopra illustrate, abbiano fornito un rilevante contributo interpretativo in merito ai processi decisionali interamente automatizzati e alle tutele spettanti alle persone fisiche dagli stessi interessati.

Il riconoscimento di un diritto alla spiegazione della decisione robotizzata, nei termini di cui si è data contezza, necessita tuttavia di una collocazione sistematica all'interno di un quadro normativo in evoluzione, che, proprio in ordine alla valutazione algoritmica del merito creditizio delle persone fisiche, sembra procedere nella medesima direzione.

Una prima testimonianza in tal senso è offerta dalla nuova direttiva 2023/2225 sul credito ai consumatori (c.d. CCD II)⁴⁹, che dedica alla valutazione del merito creditizio l'intero Capo IV, sul presupposto che una stima accurata della solvibilità del soggetto finanziato, precedente la conclusione di ogni contratto di credito, risponda ad un preciso obbligo normativo, preordinato ad evitare, nel prioritario interesse del consumatore, «pratiche irresponsabili in materia di concessione di prestiti e sovraindebitamento» (art. 18, par. 1)⁵⁰.

Ora, con riferimento alle operazioni di *algorithmic credit scoring*, l'aspetto che maggiormente rileva – oltre all'individuazione in senso restrittivo dei dati utilizzabili ai fini del giudizio di affidabilità creditizia, essendo precluso, in generale, il ricorso a dati sensibili e ad informazioni tratte dai *social network* (art. 18, par. 3)⁵¹ – concerne l'espresso riconoscimento

⁴⁹ Direttiva (UE) 2023/2225 del Parlamento europeo e del Consiglio del 18 ottobre 2023 relativa ai contratti di credito ai consumatori e che abroga la direttiva 2008/48/CE, cit. La CCD II è stata adottata con l'obiettivo di incrementare il livello di tutela dei consumatori e di agevolare la formazione di un mercato unico del credito, superando le carenze rilevate nella precedente direttiva 2008/48/CE, in considerazione della diffusione di nuove tecnologie digitali e di nuovi prodotti di credito, specie nell'ambiente *on line*. Il termine di recepimento della CCD II è fissato al 20 novembre 2025, ma le disposizioni attuative si applicheranno a decorrere dal 20 novembre 2026.

⁵⁰ È qui evidente, dunque, la valorizzazione del principio del prestito responsabile, nell'accezione del *responsible lending*, essendo il finanziatore tenuto a proteggere il consumatore dal rischio di eccessivo indebitamento (*supra*, nt. 6).

⁵¹ In particolare, la CCD II stabilisce che la valutazione di merito creditizio deve essere basata – in linea con gli *Orientamenti* dell'Autorità bancaria europea su concessione e monitoraggio dei prestiti (EBA/GL/2020/06 del 20 maggio 2020, in vigore dal 30 giugno 2021) - su informazioni pertinenti e accurate relative al reddito e alle spese del consumatore, o, più genericamente, alla sua situazione economica e finanziaria (ad es., passività o impegni), ma non può includere i dati di cui all'art. 9, par. 1, GDPR (i c.d. dati sensibili, come quelli relativi alla salute), né le informazioni tratte dai *social network*. Analogamente, è previsto che i dati di cui all'art. 9 cit. e i dati ottenuti dai *social network*, ove contenuti nelle banche dati consultabili dai creditori degli Stati membri ai fini della valutazione del merito creditizio, non possano essere trattati dai creditori medesimi, né dagli intermediari del credito (art. 19, par. 5). Nei limiti di applicazione della disciplina consumeristica, pertanto, non saranno più consentiti sistemi di valutazione automatizzata del merito creditizio che sfruttino i *soft data* ottenuti dalle reti *social* (potranno invece essere utilizzati altri tipi di *soft data*), tanto più che tali dati non soddisferebbero neanche i requisiti di attendibilità prescritti dall'art. 18, par. 3. In dottrina, tuttavia, è stato sottolineato come l'utilizzo di certe categorie di dati ai fini della valutazione del merito creditizio, seppur "scoraggiato" dal legislatore, sembrerebbe potersi ammettere, in presenza del consenso esplicito dell'interessato (E. CAPOBIANCO, *Intelligenza artificiale e rapporti bancari*, in *Rivista di diritto bancario*, 2024, n. 4, pp. 1013-4; F. TRAPANI, *La nuova direttiva*

in capo all'interessato, ove le valutazioni siano state effettuate mediante trattamento automatizzato di dati, di un diritto di ottenere dal creditore l'intervento umano; diritto che, a sua volta, si sostanzia in un vero e proprio diritto ad una spiegazione, chiara e comprensibile, della valutazione del merito creditizio (compresi la logica e i rischi derivanti dal trattamento dati, nonché la rilevanza e gli effetti sulla decisione), ulteriormente funzionale alla possibilità di chiedere un riesame della stessa, o della distinta decisione relativa alla concessione del credito (art. 18, par. 8). È di tutta evidenza, dunque, che le disposizioni in esame si pongono in linea di ideale continuità con gli artt. 15 e 22 GDPR, per come interpretati dalla Corte di Giustizia, contraddistinguendosi, nondimeno, per un contenuto più puntuale⁵²: al più generico diritto di ottenere informazioni significative, fonte delle ambiguità interpretative di cui si è discusso (*supra*, par. 4), si sostituisce, invero, quello ad una spiegazione chiara e intelligibile della decisione algoritmica, che abbia ad oggetto le principali variabili, la logica e i rischi attinenti al trattamento automatizzato applicato⁵³.

Siffatte previsioni non rappresentano peraltro un *unicum*, dovendosi tenere altresì conto di un ulteriore provvedimento normativo di relevantissimo impatto sul mercato, quale il Regolamento 2024/1689 del 13 giugno 2024 sull'intelligenza artificiale (in breve, *AI Act*)⁵⁴.

Non potendo esaminare in dettaglio i contenuti di tale Regolamento, ci limitiamo qui ad osservare che, sposando un approccio di tipo *risk-based* (basato, cioè, su una logica di proporzionalità delle regole all'intensità e alla portata dei rischi che possono essere generati dai diversi sistemi di IA), l'*AI Act* circonda di particolari cautele i sistemi c.d. «ad alto rischio»⁵⁵, tra i quali vengono annoverati anche quelli per la valutazione automatizzata del merito creditizio o dell'affidabilità creditizia delle persone fisiche: questi, invero, determinano l'accesso delle persone alle risorse finanziarie o a servizi essenziali (alloggio, servizi di telecomunicazione, elettricità) e, se progettati o utilizzati in modo non corretto, possono produrre effetti deleteri, come l'esclusione finanziaria e la discriminazione di singoli individui o di gruppi⁵⁶.

Consapevole, dunque, della pericolosità di siffatti sistemi, il legislatore prevede una serie di misure volte a neutralizzarne, per quanto possibile, gli effetti nocivi, richiedendo, a monte del loro impiego, l'uso di dati che soddisfino specifici requisiti di qualità e il ricorso ad una struttura quanto più possibile comprensibile e trasparente, unitamente all'adozione di apposite misure di prevenzione e gestione dei rischi; a valle, l'intelligibilità dell'*output* del sistema, che il *deployer* (chi utilizza il sistema di IA sotto la propria autorità, al di fuori di

2023/2225/UE *sul credito al consumo: note in tema di educazione finanziaria, merito di credito e servizi di consulenza sul debito*, in *Le nuove leggi civili commentate*, 2024, n. 3, p. 779; *contra*, C. A. VALENZA, *op. cit.*, p. 277, secondo il quale il trattamento di certi dati è vietato, a prescindere da un eventuale consenso del consumatore; G. FALCONE, *Prime riflessioni sulla Direttiva CCD II: le informazioni e la valutazione del merito creditizio*, in *Diritto della banca e del mercato finanziario*, 2024, n. 4, pp. 625-7).

⁵² Lo evidenzia M. PROTO, *Contrattazione tecnobancaria e gestione dei dati*, cit., c. 122-3, sottolineando che il consumatore avrebbe diritto ad ottenere «non un generico contraddittorio come previsto dall'art. 22, par. 3, GDPR, ma la ripetizione della valutazione senza l'ausilio della macchina; al contempo, non solo «informazioni significative sulla logica utilizzata», come stabilito dall'art. 15, par. 1, lett. h), Gdpr, ma una preventiva, completa e non ambigua, illustrazione di essa».

⁵³ Considerando n. 56 della CCD II.

⁵⁴ Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale, in *GUUE L* del 12 luglio 2024. Com'è noto, l'*AI Act* si applicherà integralmente a far data dal 2 agosto 2026.

⁵⁵ Nell'impianto dell'*AI Act*, i sistemi ad alto rischio sono quelli in grado di produrre un significativo impatto negativo sulla salute, sulla sicurezza o sui diritti fondamentali delle persone.

⁵⁶ Considerando n. 58.

un'attività personale non professionale) deve poter correttamente interpretare (anche per stabilire se avvalersene o meno, o se capovolgerlo), nonché saper spiegare al soggetto interessato da una decisione adottata sulla scorta di quell'*output*⁵⁷. Proprio in questa prospettiva, pertanto, è stato anche in questo caso sancito un diritto alla spiegazione del singolo processo decisionale, avente ad oggetto sia il ruolo del sistema di IA nel contesto della procedura, che i principali elementi dello specifico provvedimento adottato (art. 86).

Ancora una volta, dunque, il legislatore europeo insiste sulla necessità di trasparenza e decifrabilità delle decisioni automatizzate (quanto meno ove le stesse siano frutto di un sistema ad alto rischio, come quelli di *algorithmic credit scoring*), ribadendo che le stesse devono essere rese comprensibili al diretto interessato, per consentirgli un utile esercizio dei diritti e dei mezzi di tutela allo stesso riconosciuti dal diritto dell'Unione⁵⁸.

In altri termini, nella visione dichiaratamente antropocentrica che permea le nuove norme sull'intelligenza artificiale, non può in alcun modo ammettersi che l'essere umano resti assoggettato al potere decisionale assoluto di un apparato tecnologico in grado di incidere in modo così pervasivo sulla vita privata, economica e sociale delle persone. L'adozione di regole di trasparenza in senso lato, al cui interno possono a pieno titolo annoverarsi anche quelle relative al diritto alla spiegazione della decisione algoritmica - affermato in più sedi - risponde dunque ad una precisa *ratio*: impedire che la persona divenga materia passiva di decisioni artificiali non comprensibili, e come tali non contestabili, il cui effetto sarebbe, in ultima analisi, una inaccettabile violazione del fondamentale valore della dignità umana⁵⁹.

7. *Credit scoring* e *social scoring*. La ricerca di una linea di confine

Fermo restando quanto sopra, l'adozione dell'*AI Act* sollecita comunque un'ulteriore riflessione, con riguardo al tema del *credit scoring* automatizzato.

Le regole di nuovo conio, invero, ammettono, come già detto, i sistemi (ad alto rischio) per la valutazione automatizzata del merito creditizio o dell'affidabilità creditizia delle

⁵⁷ In quest'ottica vanno lette le disposizioni dell'*AI Act* che prevedono, ad es., la predisposizione di un sistema di individuazione, valutazione e gestione dei rischi, volto a permetterne l'eliminazione o l'attenuazione (art. 9); l'impiego di *set* di dati di addestramento, convalida e prova pertinenti, completi e quanto più possibile esenti da errori, soggetti a misure di *governance* idonee a prevenire o attenuare eventuali distorsioni (art. 10); un grado di trasparenza tale da consentire al *deployer* di interpretare l'*output* del sistema e di utilizzarlo adeguatamente (art. 13); la sorveglianza da parte di una persona fisica, che possa monitorare il funzionamento del sistema, interpretarne correttamente l'*output* e intervenire, all'occorrenza, decidendo di ignorarlo, annullarlo o ribaltarlo, o di arrestare il sistema in condizioni di sicurezza (art. 14); la valutazione dell'impatto del sistema sui diritti fondamentali della persona (c.d. FRIA), in chiave di prevenzione di possibili abusi o discriminazioni (art. 27); infine, il diritto dell'interessato, ove non previsto da altre norme del diritto dell'Unione, di ottenere una spiegazione chiara e significativa del ruolo svolto dal sistema ad alto rischio in una specifica procedura decisionale e dei principali elementi della decisione adottata (art. 86).

⁵⁸ Non è possibile approfondire, in questa sede, il rapporto fra le previsioni dell'*AI Act* e del GDPR (in arg., P. FALLETTA, A. MARSANO, *Intelligenza artificiale e protezione dei dati personali: il rapporto tra Regolamento europeo sull'intelligenza artificiale e GDPR*, in *Rivista italiana di informatica e diritto*, 2024, p. 119 ss.), bastando qui evidenziare che il diritto alla spiegazione ex art. 86 *AI Act* si applica solo in via residuale, ove non previsto già da altre disposizioni, come ad es. l'art. 22 GDPR. Sui problemi derivanti dalla coesistenza di due distinti diritti alla spiegazione delle decisioni algoritmiche, non applicabili simultaneamente, v. A. G. GRASSO, *Decisioni automatizzate e merito creditizio*, cit., p. 742.

⁵⁹ R. MESSINETTI, *La tutela della persona umana versus l'intelligenza artificiale*, cit., p. 892.

persone fisiche, ma vietano, per altro verso, l'immissione sul mercato, la messa in servizio o l'uso dei sistemi di intelligenza artificiale che permettono l'attribuzione di un punteggio sociale alle persone fisiche sulla base di dati riguardanti il comportamento sociale o di caratteristiche personali o della personalità note, inferite o previste, là dove tale punteggio comporti effetti sfavorevoli o pregiudizievoli in contesti diversi da quelli in cui i dati sono stati generati o raccolti, o effetti sfavorevoli o pregiudizievoli non giustificati o non proporzionati rispetto alla gravità del comportamento sociale rilevato (art. 5, «Pratiche di IA vietate»). Tali sistemi – che trovano il più noto esempio nel famigerato *social credit system* implementato nella Repubblica Popolare Cinese – sollevano delicatissime questioni sul piano sia etico che giuridico, giacché la pretesa di “recensire” gli esseri umani, misurandone la reputazione e l'affidabilità in base alle azioni tenute in un determinato contesto sociale, può seriamente «ledere il diritto alla dignità e alla non discriminazione e i valori di uguaglianza e giustizia»⁶⁰.

Ebbene, l'aspetto che qui rileva è rappresentato dalle evidenti similitudini tra i suddetti sistemi di *social scoring* e quelli di *credit scoring* algoritmico: attraverso l'elaborazione di dati relativi alla condotta o alle caratteristiche di una persona, infatti, entrambi forniscono dei punteggi aventi una spiccata valenza “reputazionale”⁶¹, dai quali possono discendere anche conseguenze pregiudizievoli per l'interessato.

Per vero, in dottrina si è cercato di mettere in luce le differenze tra i *software* algoritmici reputazionali e quelli di *credit scoring*, evidenziando, in particolare, come i primi si basino sulla valutazione del comportamento passato degli individui, per calcolarne la credibilità sul piano sociale, ma sarebbero per tale ragione privi – malgrado una diffusa (ed erronea) percezione di segno contrario – di quelle capacità predittive, costruite su stime probabilistiche, tipiche dei secondi⁶². In altre parole, mentre un sistema di *social scoring* è impostato essenzialmente sull'analisi di dati storici (e sulla base di questi ultimi può stimare l'affidabilità di un soggetto sotto diversi aspetti, compreso, se del caso, quello finanziario), i sistemi di *credit scoring* algoritmico si fondano sul presupposto che i comportamenti degli individui appartenenti a categorie omogenee tendono a ripetersi in modo uniforme nel tempo, rendendo così prevedibile il futuro verificarsi, con maggior o minor grado di probabilità, di certi eventi (nella specie, l'inadempimento delle obbligazioni nascenti da un finanziamento, richiesto da un soggetto incasellato in un certo profilo). Si tratterebbe, quindi, di situazioni non assimilabili, quanto meno sotto il profilo dell'operatività tecnica e delle capacità previsionali.

Al di là di tali differenze, tuttavia, il timore è che, nella prassi, la linea di confine tra i due tipi di sistemi possa sensibilmente attenuarsi e che, sotto l'etichetta di sistemi di valutazione dell'affidabilità creditizia delle persone (legittimi, ancorché ritenuti ad alto

⁶⁰ Considerando n. 31. Una descrizione dell'inquietante *social credit system* cinese è effettuata da G. CERRINA FERONI, *Intelligenza artificiale e sistemi di scoring sociale*, cit., p. 4, che ne illustra gli effetti negativi per i cittadini ritenuti poco affidabili (per via di punteggi bassi), in termini di divieto di acquistare voli domestici, di impossibilità di iscrivere i figli in determinate scuole, di limitazioni nella fruizione di determinati servizi (ad es., connessione *internet* o carte di credito), di esclusione da certe posizioni lavorative e persino di inserimento in una sorta di *black list* pubblica, dal quale potrebbe conseguire l'applicazione di una o più penalità.

⁶¹ Anche il *credit scoring*, invero, si traduce in un punteggio che pretende di misurare un rilevante aspetto sociale della persona interessata, quale la sua affidabilità sul piano finanziario e, quindi, la sua reputazione di buon/cattivo pagatore.

⁶² E. FALLETTI, *op. cit.*, p. 127, ove si rileva criticamente la «presunta capacità attribuita ad un algoritmo reputazionale di far coincidere la reputazione (ovvero ciò che è inerente ad una persona rispetto al suo passato) con ciò che potrebbe succedere in futuro, ma che nonostante la buona volontà e la programmazione, nessuno è in grado di conoscere (cioè una predizione)».

rischio), finiscano col celarsi veri e propri sistemi di *social scoring* di tipo reputazionale (in linea di principio vietati)⁶³.

Il dubbio si pone, invero, proprio con riferimento a casi come quello esaminato dalla Corte di Giustizia nella causa C-203/22 (*De'B*): il rifiuto di stipulare un contratto di telefonia per un corrispettivo davvero irrisorio (10 euro mensili) lascia presumere, infatti, o che sia stato adottato un sistema di valutazione dell'affidabilità creditizia del consumatore talmente sofisticato e attendibile da riuscire a prevedere persino il mancato pagamento di importi minimi, o che sia stato surrettiziamente posto in essere un metodo di valutazione sociale dell'individuo, basato sull'analisi di dati raccolti nei più svariati contesti e delle pregresse condotte dell'interessato. In simili situazioni, invero, il risultato del sistema di profilazione del consumatore sembrerebbe esprimere, più che il reale grado di solvibilità di quest'ultimo (difficilmente calcolabile rispetto all'adempimento di obblighi particolarmente modesti), un più ampio giudizio di tipo reputazionale, ricavato dall'analisi dei comportamenti tenuti dal soggetto censito, anche nell'ambiente digitale, o delle sue caratteristiche personali⁶⁴.

Il rischio da scongiurare, pertanto, è che il diffuso impiego di sistemi di valutazione dell'affidabilità creditizia delle persone fisiche, anche al di fuori dei contesti nei quali tale pratica risulti effettivamente necessaria (come nel settore finanziario in senso lato, ove il calcolo del tasso di solvibilità del debitore risponde a precisi obblighi normativi, volti a soddisfare interessi anche pubblicistici: *supra*, par. 1), consenta di fare rientrare dalla finestra una pratica (il *social scoring*) che, per ragioni legate alla sua intrinseca pericolosità, il legislatore ha ritenuto di lasciare fuori dalla porta⁶⁵.

8. Considerazioni finali. La trasparenza algoritmica, tra tecniche di eXplainable AI e strumenti di enforcement

L'elemento che emerge con lampante chiarezza, dalle riflessioni fin qui sviluppate in materia di *credit scoring* algoritmico, è il ruolo che il diritto eurounitario affida al principio della trasparenza, ai fini di tutela della persona scrutinata. Tale principio, invero, rappresenta il *trait d'union* fra i provvedimenti normativi sopra analizzati (GDPR, CCD II e *AI Act*), ciascuno dei quali sembra preordinato a conseguire, al netto delle differenze nelle rispettive formulazioni letterali, il medesimo obiettivo: l'intelligibilità delle decisioni algoritmiche riguardanti l'affidabilità creditizia delle persone fisiche, a presidio della quale è posto quel

⁶³ Per un primo spunto in questo senso v. F. MATTASSOGLIO, *Innovazione tecnologica*, cit., p. 166, secondo la quale le attuali tecnologie, basate sull'uso dei *big data*, sospingono sempre più verso l'introduzione di «un vero e proprio sistema di *social credit scoring* di tipo virtuale»; E. DI CARPEGNA BRIVIO, *Il Reputation scoring e la quantificazione del valore sociale*, in *Federalismi.it*, 2022, pp. 126-7, la quale rileva la tendenza alla trasformazione dello *scoring* creditizio in *scoring* reputazionale, dal momento che l'analisi di dati eterogenei, non soltanto finanziari (propensioni al consumo, gusti, ricerche *on line*, attività sui *social media*, frequentazioni, ambiente lavorativo dell'interessato, ecc.), consente di valutare la generale affidabilità sociale del soggetto valutato, con riflessi anche in termini di accesso al credito.

⁶⁴ Viene alla memoria il noto caso della riduzione di un fido concesso ad un cittadino afroamericano con un'ottima storia creditizia, eppure penalizzato da acquisti con carta di credito effettuati presso esercizi commerciali frequentati da clientela poco affidabile (l'episodio, relativo ad una discutibile forma di c.d. *creditworthiness by association*, è citato, da ultimo, da G. MATTARELLA, *op. cit.*, p. 75).

⁶⁵ S. PESUCCI, *Critica etica al sistema di credit scoring automatizzato*, 30 ottobre 2024, p. 18, reperibile su www.ilcaso.it.

“diritto alla spiegazione” che la Corte di Giustizia, nelle sue recenti pronunce, ha provveduto a meglio delineare nella sua dimensione pratica e applicativa.

La trasparenza, nell’accezione sopra indicata (intesa cioè, a riprova della polisemia del termine, non nel classico senso di conoscibilità e chiarezza del contenuto negoziale, ma, piuttosto, come spiegabilità del procedimento decisionale governato dalla macchina), assume dunque la valenza di un principio-obiettivo, ricollegabile, peraltro, ad ulteriori principi altrettanto rilevanti, come quelli di “non esclusività algoritmica” o di “sorveglianza umana”, comunemente sintetizzati nell’espressione gergale *human in the loop*. L’obbligo di assicurare l’intervento umano, ovvero, in termini più ampi, l’assoggettamento del sistema decisionale algoritmico al controllo dell’uomo, può realizzarsi in modo efficace, infatti, solo se la persona cui è affidato il potere di azione o di sorveglianza sul sistema stesso è in grado di comprenderne il funzionamento, di rilevare eventuali disfunzioni o anomalie e, soprattutto, di interpretarne correttamente l’*output*, anche nell’ottica di una modifica o di un annullamento della decisione automatizzata. Tutte situazioni che presuppongono, con ogni evidenza, la trasparenza sia del processo decisionale che dei risultati che ne costituiscono l’esito.

Si è già avuto modo di osservare, nondimeno, come la trasparenza possa rivelarsi, rispetto a sistemi decisionali sofisticati come quelli di *credit scoring* algoritmico, un obiettivo ambizioso e fallace al tempo stesso, di difficile realizzazione pratica. Ciò non soltanto a causa della già rilevata difficoltà di ricostruire il percorso logico seguito dalla macchina (il riferimento è ancora una volta ai sistemi di *machine learning*), ma anche per via di limiti di carattere normativo, che impediscono la *disclosure* di certi dati per ragioni legate specialmente alla tutela della proprietà intellettuale e del *know-how*.

Ora, sotto il primo aspetto, la via da seguire pare obbligata: fare ricorso ad evolute tecniche di c.d. *eXplainable AI (XAI)*⁶⁶, ancora oggi in fase di sviluppo, che permettano di coniugare affinate capacità predittive e spiegabilità della decisione algoritmica (superando, in tal modo, il *trade-off* tra trasparenza e accuratezza delle previsioni, tipico dei sistemi di *machine learning*)⁶⁷. In caso contrario, del resto, sistemi ad alto rischio come quelli di *credit scoring* non dovrebbero neanche trovare ingresso sul mercato, per difformità dai requisiti di legge⁶⁸.

Quanto al secondo profilo, invece, si tratta di raggiungere un punto di equilibrio ottimale tra la tutela di interessi contrapposti, posto che, come già osservato, il diritto alla protezione dei dati personali non rappresenta una prerogativa assoluta⁶⁹, ma va temperato, nel rispetto di una logica di proporzionalità, con altri diritti che ricevono protezione giuridica, come il segreto commerciale (art. 2, dir. 2016/943) o la proprietà intellettuale (espressamente tutelata dall’art. 17 Carta dei diritti fondamentali dell’Unione europea). L’ostensione di dati coperti da diritti di privativa (in specie, il codice sorgente del *software*) dovrebbe essere ammessa, pertanto, solo laddove il sacrificio di questi ultimi risponda ad un’effettiva e concreta esigenza di tutela dell’interessato, non altrimenti realizzabile con strumenti meno invasivi, giacché, in caso contrario, la lesione dei diritti altrui

⁶⁶ A. ABUSITTA, M.Q. LI, B.C.M. FUNG, *Survey on Explainable Ai: Techniques, Challenges and Open Issues*, 2024 (<https://ssrn.com/abstract=4710037>); M. SCHMITT, *Explainable Automated Machine Learning for Credit Decisions: Enhancing Human Artificial Intelligence Collaboration in Financial Engineering*, 2024 (<https://ssrn.com/abstract=4716847>).

⁶⁷ AA. VV., *Intelligenza artificiale nel credit scoring*, cit., p. 15; E. FALLETTI, *op. cit.*, p. 115-6, ove i riferimenti a studi che comprovano la realizzabilità sul piano tecnico, anche nell’ambito del *credit scoring*, di modelli totalmente interpretabili.

⁶⁸ P. GAGGERO, C. A. VALENZA, *op. cit.*, p. 858; A. M. G. GALATO, *Credit scoring e trasparenza algoritmica*, cit., pp. 40-1.

⁶⁹ Considerando n. 4 e n. 63 GDPR; art. 23, par. 1, GDPR.

risulterebbe non necessaria e ingiustificata. Occorre, quindi, rifuggire da soluzioni astratte e aprioristiche, che pretendano di collocare in un ordine gerarchico prestabilito gli interessi in gioco, e valutare invece caso per caso le ragioni fatte valere dall'interessato in sede di esercizio del diritto di accesso, rigettando, ad es., istanze aventi contenuto generico o meramente "esplorativo", o prive di indicazione dei motivi che possano effettivamente giustificare l'ostensione di dati riservati⁷⁰. Il bilanciamento tra interessi collidenti, pertanto, sembra doversi operare alla luce di uno specifico criterio ordinatore: la comunicazione di dati e informazioni protette da segreto aziendale o dal diritto d'autore sul *software* utilizzato deve restare confinata a situazioni eccezionali e di stretta necessità, tanto più che, nella generalità dei casi, la trasmissione di informazioni di natura tecnico-informatica, scarsamente comprensibili da parte di soggetti privi di competenze specialistiche, si rivelerebbe poco utile all'esercizio delle ragioni dell'interessato (e, dunque, anche sotto tale aspetto sproporzionata rispetto al sacrificio dei diritti altrui). Né può ritenersi che il risultato di tale bilanciamento, imposto dal diritto dell'Unione, venga stabilito in modo definitivo dal singolo Stato membro, come più volte ribadito dalla giurisprudenza della Corte⁷¹: solo un futuro intervento del legislatore europeo, in sostanza, potrebbe attribuire priorità ad un diritto rispetto ad un altro, operando una precisa scelta politica, allo stato inattuata.

Ci sia consentita ancora qualche rapida riflessione, prima di addivenire alle conclusioni.

Una prima notazione attiene al piano dei rimedi esperibili a fronte di una violazione del diritto alla spiegazione della decisione algoritmica, configurabile sia nel caso in cui nessuna informazione venga trasmessa all'interessato, sia nel caso in cui le informazioni trasmesse appaiano insufficienti a ricostruire la logica sottesa al singolo provvedimento. Inutile precisare che il punto è di notevole rilevanza, poiché solo in presenza di un adeguato apparato rimediale il diritto da tutelare acquista piena effettività⁷². Nel caso del *credit scoring* algoritmico, tuttavia, il sistema rimediale delineato dall'attuale *framework* legislativo non appare particolarmente efficace, rivelando un punto debole della normativa di riferimento.

Nel GDPR, invero, la violazione del diritto di accesso *ex art. 15* non appare corredata dalla previsione di alcun rimedio, che solo in sede dottrina è stato individuato nell'illiceità del processo decisionale automatizzato (in quanto non assistito dalle necessarie garanzie per l'interessato) e nella conseguente inefficacia della decisione (opaca) nella sfera giuridica del destinatario della stessa⁷³. Bisognerebbe domandarsi, tuttavia, fino a che punto tale soluzione soddisfi l'interesse della persona profilata, la cui aspirazione, nella specie, è quella di ottenere un finanziamento, confidando in una valutazione positiva della propria solvibilità. Anche ipotizzando, infatti, che, in assenza delle prescritte «informazioni significative», lo *scoring*

⁷⁰ Tale criterio riecheggia, in qualche modo, quello attraverso il quale il Consiglio di Stato ha risolto il rapporto tra diritto di accesso difensivo e tutela della riservatezza (nella specie, finanziaria ed economica), asserendo che l'ostensione di un documento non può essere legittimata da un generico riferimento a non meglio precisate esigenze probatorie e difensive, ma presuppone «un rigoroso, motivato, vaglio sul nesso di strumentalità necessaria tra la documentazione richiesta e la situazione finale che l'istante intende curare o tutelare» (Cons. Stato, ad. plen., 18 marzo 2021, n. 4).

⁷¹ CGUE, 7 dicembre 2023, C-634/21 (*Schufa*), punto 70; CGUE, 19 ottobre 2016 C-582/14 (*Breyer*), punto 62.

⁷² Giova chiarire che l'interesse ad ottenere la spiegazione dello *scoring* algoritmico sussiste in caso di attribuzione di un punteggio sia negativo che positivo. Nella prima ipotesi, invero, il soggetto censito ha interesse a comprendere la decisione algoritmica che gli ha precluso l'erogazione del credito, per poterla eventualmente contestare; nella seconda, invece, potrebbe essere interessato a rilevare l'erroneità di uno *scoring* positivo, cui abbia fatto seguito la concessione di un finanziamento insostenibile, in violazione del principio del prestito responsabile. Il tema finisce per saldarsi, dunque, a quello, ampiamente dibattuto in dottrina, della responsabilità dell'intermediario per erronea o carente valutazione del merito creditizio (v. *supra*, nt. 8).

⁷³ R. MESSINETTI, *La tutela della persona umana versus l'intelligenza artificiale*, cit., pp. 890-1.

creditizio formulato in termini negativi non produca effetti nei confronti del soggetto censito, questi non riuscirebbe comunque ad avere accesso al credito, in difetto di una valutazione di diverso tenore (né, tanto meno, disporrebbe di strumenti giuridici atti a costringere l'intermediario ad erogare il prestito, essendo noto che il nostro ordinamento non contempla alcun obbligo di fare credito, neanche nei confronti di soggetto meritevole⁷⁴).

Analoghe considerazioni valgono, poi, per l'*AI Act*, che non disciplina espressamente le conseguenze del mancato rispetto dell'obbligo di spiegazione del singolo processo decisionale, sancito, sia pure in forma residuale rispetto a previsioni di analogo contenuto, dall'art. 86. Da questo punto di vista, dunque, tutele maggiormente incisive sembrano offerte dall'art. 18 CCD II, che prevede, quanto meno, il diritto del consumatore ad ottenere un riesame, da parte di un essere umano, della valutazione del merito creditizio operata mediante trattamento automatizzato di dati e della decisione relativa alla (mancata) concessione del finanziamento (laddove il GDPR riconosce soltanto, in termini più generici, il «diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione»⁷⁵). Anche in questo caso, tuttavia, il limite è rappresentato dall'impossibilità di obbligare l'intermediario alla stipula del contratto di credito, trattandosi, come prima accennato, di valutazione non coercibile, che pertiene all'ambito delle scelte imprenditoriali discrezionali del finanziatore⁷⁷; al più, dunque, potrebbe ipotizzarsi, in caso di mancata conclusione dell'accordo, una tutela di tipo risarcitorio legata alla presenza di «anomalie nel processo decisionale», sempre che ricorrano i necessari presupposti⁷⁸.

Si rammenti, peraltro, che l'applicazione di rimedi privatistici è complementare al ricorso a misure di *public enforcement* (ad es., irrogazione di sanzioni pecuniarie o amministrative), da parte delle autorità cui sono attribuiti, nei rispettivi ambiti di competenza, poteri di intervento e sanzionatori (il riferimento è agli enti preposti alla tutela dei dati personali, al controllo sul mercato del credito, alla tutela dei consumatori e alla vigilanza del mercato ai fini del regolamento sull'intelligenza artificiale⁷⁹). Si giunge così ad individuare un ulteriore profilo di criticità, sul quale merita appuntare l'attenzione: la necessità di un coordinamento sistematico fra più norme concorrenti, in punto di valutazioni automatizzate

⁷⁴ Diffusamente, sul punto, G. LIBERATI BUCCIANI, *op. cit.*, p. 81 ss. *Contra*, ma in posizione minoritaria, F. SARTORI, *Deviazioni del bancario e dissociazione dei formanti: a proposito del diritto al credito*, in *Giustizia civile*, 2015, p. 600.

⁷⁵ V. *supra*, nt. 52.

⁷⁶ Resta salva, per altro verso, l'introduzione, da parte degli Stati membri, di sanzioni effettive, proporzionate e dissuasive, da applicare in caso di violazione delle disposizioni attuative della direttiva (art. 44). L'auspicio, pertanto, è che il legislatore italiano possa disciplinare, in sede di attuazioni della CCD II, anche le conseguenze della violazione dell'obbligo di spiegazione dello *scoring* automatizzato.

⁷⁷ Con il temperamento rappresentato, tuttavia, dal rispetto degli obblighi di correttezza e buona fede, cui deve essere improntata anche la fase delle trattative, sicché l'intermediario è tenuto a fornire al cliente tempestivi e adeguati chiarimenti circa il rifiuto di concedere il prestito, in mancanza dei quali si esporrebbe a responsabilità precontrattuale (ABF, Coll. Coord., dec. n. 6182/13; D. POLETTI, *La responsabilità della banca per diniego del credito nelle decisioni dell'Arbitro Bancario Finanziario*, in *Responsabilità civile e previdenza*, 2014, n. 5, p. 1423).

⁷⁸ G. FALCONE, *Prime riflessioni sulla Direttiva CCD II*, cit., p. 622-3.

⁷⁹ Ai sensi dell'art. 74 *AI Act*, per i sistemi di IA ad alto rischio immessi sul mercato, messi in servizio o usati da istituti finanziari disciplinati dal diritto dell'Unione in materia di servizi finanziari, l'autorità di vigilanza del mercato coincide, di norma, con l'autorità nazionale di vigilanza su tali enti, nella misura in cui l'immissione sul mercato, la messa in servizio o l'uso del sistema di IA siano direttamente collegati alla fornitura di tali servizi finanziari (par. 6).

del merito creditizio⁸⁰, cui si accompagna l'inevitabile moltiplicarsi dei soggetti pubblici chiamati ad intervenire in caso di violazioni di legge (dove l'esigenza di un raccordo fra autorità, improntato ad uno spirito di leale collaborazione, che possa, da un lato, limitare il rischio di sovrapposizioni e conflitti, dall'altro, contribuire ad una razionalizzazione del quadro normativo complessivo, nell'interesse di tutti i soggetti coinvolti⁸¹).

La sensazione è che, in un contesto regolatorio così articolato, e per alcuni aspetti ancora non sufficientemente definito, le disposizioni in materia di *credit scoring* algoritmico (e, più in generale, l'intera disciplina delle decisioni automatizzate) non riescano a centrare del tutto l'obiettivo per il quale sono state emanate: assicurare la trasparenza algoritmica, in funzione di una piena ed effettiva tutela della persona rispetto al rischio di incondizionata soggezione al potere decisionale della macchina. La difficoltà di garantire, sul piano concreto, l'intelligibilità dei processi decisionali automatizzati; il conflitto fra obblighi informativi e protezione della proprietà industriale e del segreto commerciale; la coesistenza di più norme (e di più strumenti di *private* e *public enforcement*) applicabili alla medesima fattispecie, sono elementi che minano la coerenza e la credibilità del *framework* normativo e, con essa, la fiducia del pubblico verso le nuove tecnologie, frenandone lo sviluppo e rischiando di alimentare un vasto contenzioso, del quale si sono avute già le prime avvisaglie. Su un punto, tuttavia, non ci si può permettere di arretrare: lo sviluppo tecnologico, per quanto foriero di maggiore rapidità, accuratezza ed efficienza nella prestazione dei servizi finanziari, non dovrà mai prevalere sulla dignità della persona umana e sulla tutela dei suoi diritti fondamentali.

⁸⁰ Si pensi ancora una volta, ad es., al rapporto tra il diritto alla spiegazione *ex art.* 86 *AI Act* e quello previsto dall'art. 22 GDPR (*supra*, nt. 58); o ancora, al distinto ambito di applicazione dell'art. 22 GDPR (riferito al trattamento interamente automatizzato di dati personali non meglio specificati) e dell'art. 18, par. 8, CCD II (relativo al trattamento, anche non del tutto automatizzato, di dati personali relativi alla situazione economico-finanziaria del consumatore, purché non tratti dai *social network*).

⁸¹ Emblematico, al riguardo, l'art. 99 *AI Act*, che rimette agli Stati membri il compito di stabilire sanzioni effettive, proporzionate e dissuasive, per i casi di violazione del Regolamento da parte degli operatori, precisando poi che, nel decidere se infliggere una sanzione amministrativa pecuniaria e nel determinarne l'importo debba tenersi conto, tra l'altro, dell'eventuale applicazione allo stesso operatore, da parte di altre autorità di vigilanza del mercato, di sanzioni amministrative pecuniarie per la stessa violazione o per violazioni di altre disposizioni del diritto dell'Unione o nazionale, qualora tali violazioni derivino dalla stessa attività o omissione che costituisce una violazione del regolamento.



CRISTINA GRIECO*

DAL MICROTARGETING POLITICO AI DEEPPAKE: COME LE NUOVE TECNOLOGIE STANNO RISCRIVENDO LA DEMOCRAZIA IN EUROPA

SOMMARIO: 1. Introduzione. – 2. I rischi del *microtargeting* politico e di un *gerrymandering* digitale: l'azione europea a difesa della democrazia – 3. Il quadro giuridico europeo di riferimento: DSA, PAR e *AI Act* alla luce della Carta dei diritti fondamentali – 4. Conclusioni.

1. Introduzione

L'impiego sempre più ampio e pervasivo delle nuove tecnologie¹ sta sollevando interrogativi inediti capaci di mettere in discussione la tenuta dei sistemi democratici². L'utilizzo di algoritmi di profilazione e di sofisticati strumenti di *microtargeting* nella comunicazione politica, la diffusione di contenuti manipolati, i *deepfake* e la creazione di *filter bubble* nei *social media* sono in grado di incidere sull'opinione collettiva e di amplificare fenomeni di polarizzazione del dibattito pubblico e di disinformazione³.

In un simile contesto, nel consueto bilanciamento dei tradizionali poteri statali – legislativo, esecutivo e giudiziario – all'interno del quale va altresì considerato il ruolo dell'informazione, ormai comunemente considerata alla stregua di un quarto potere, si è imposto un ulteriore polo di influenza, quello computazionale⁴, rappresentato dalle tecnologie digitali e dai grandi attori del settore. In effetti, per l'entità dell'influenza che le

* Ricercatore in *Tenure Track* in Diritto dell'Unione europea, Università degli Studi di Macerata, Dipartimento di Giurisprudenza, Dipartimento di Eccellenza 2023-2027.

¹ L. FLORIDI, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford, 2014.

² O. POLLICINO, P. DUNN, *Intelligenza artificiale e democrazia. Opportunità e rischi di disinformazione e discriminazione*, Milano, 2022, p. 1 ss.

³ La disinformazione è la diffusione intenzionale di informazioni false o fuorvianti per influenzare il dibattito pubblico o i processi elettorali, anche tramite contenuti formalmente leciti. Si distingue dalla misinformazione, che invece è la condivisione involontaria di notizie errate. Si veda la definizione proposta dalla Commissione europea all'interno della Comunicazione COM/2020/790 final del 3 dicembre 2020, sul *Piano d'azione per la democrazia europea*, p. 20. Si veda altresì European Commission, *A multi-dimensional approach to disinformation. Report of the independent High level Group on fake news and online disinformation*, Luxembourg, 2018, p. 10.

⁴ Sul punto M. DURANTE, *Potere computazionale: l'impatto delle ICT su diritto, società, sapere*, Milano, 2019.

Tech Giants sono in grado di esercitare risultano sempre più assimilabili a poteri parastatali, delle vere e proprie oligarchie del digitale, in grado di condizionare in modo significativo anche il dibattito politico. Per comprendere la portata del fenomeno, basti considerare che la capitalizzazione di mercato delle sole *Google, Microsoft, Apple, Meta, Alphabet* e *Tesla* sfiora la metà del PIL statunitense e si avvicina all'intero PIL dell'Unione europea.

La fiducia nella *governance* digitale appare dunque strettamente correlata alla capacità del legislatore sovranazionale di assicurare che sul mercato vengano immessi sistemi e servizi affidabili, verificabili, conformi a elevati *standard* di sicurezza⁵ all'interno dei quali venga promosso un approccio *digital by default, inclusive by design*.

In una simile prospettiva, il processo di integrazione europea appare sempre più legato ad una dimensione nuova, quella della coesione digitale, che viene ad affiancarsi ai tradizionali ambiti economico, sociale e territoriale.

Alla luce di quanto sopra, il presente contributo si propone di analizzare l'attuale quadro giuridico europeo al fine di valutarne l'idoneità a salvaguardare la democrazia e i cittadini da fenomeni come la disinformazione, il *microtargeting* politico e la manipolazione online.

2. I rischi del microtargeting politico e di un gerrymandering digitale: l'azione europea a difesa della democrazia

Nell'attuale contesto del cosiddetto "capitalismo della sorveglianza"⁶, il crescente impiego delle tecnologie digitali e dei *social media* per finalità manipolative suscita rilevanti preoccupazioni sotto il profilo della tutela delle libertà individuali e di stampa. L'utilizzo non neutrale di tali strumenti rischia, infatti, di compromettere la fiducia dei cittadini nelle istituzioni democratiche, con potenziali ricadute sull'integrità del dibattito pubblico.

L'anno 2024, caratterizzato da un numero eccezionalmente elevato di tornate elettorali, ha rappresentato un punto di svolta negli equilibri geopolitici globali⁷. In un simile scenario, si è delineato con chiarezza il ruolo ormai imprescindibile dei *social media* quali strumenti di propaganda e mobilitazione elettorale, parallelamente all'emersione di nuovi e gravi rischi connessi alla diffusione di disinformazione, alla produzione di *deepfake* e alle interferenze di soggetti esterni. Tali fenomeni sono stati qualificati come una vera e propria "minaccia ibrida" per la democrazia, idonea a compromettere sia l'integrità delle consultazioni elettorali, sia il corretto funzionamento delle istituzioni rappresentative⁸.

⁵ Sul tema nuove tecnologie e diritti fondamentali si vedano P. DE PASQUALE, *Verso una Carta dei diritti digitali (fondamentali) dell'Unione europea?*, in *Il Diritto dell'Unione europea – Osservatorio*, 17 marzo 2022; A. ADINOLFI, *Evoluzione tecnologica e diritti fondamentali: qualche considerazione sulle attuali strategie normative dell'Unione*, in *I Post di AISDUE*, 2023, p. 323 ss.; C. GRIECO, *Intelligenza artificiale e tutela degli utenti nel diritto dell'Unione europea*, Napoli, 2023.

⁶ S. ZUBOFF, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, London, 2019.

⁷ O. POLLICINO, P. DUNN, *Disinformazione e intelligenza artificiale nell'anno delle global elections: rischi (ed opportunità)*, in *federalismi.it*, 29 maggio 2024, p. 12; G. M. RUOTOLO, *Nell'anno delle elezioni hanno tutti ragione. Alcune considerazioni sul ruolo di Diritto internazionale ed UE nel contrasto alla disinformazione*, in *SIDIblog*, 5 aprile 2024.

⁸ Nella comunicazione della Commissione C/2024/3014 del 26 aprile 2024, *Orientamenti della Commissione per i fornitori di piattaforme online di dimensioni molto grandi e di motori di ricerca online di dimensioni molto grandi sull'attenuazione dei rischi sistemici per i processi elettorali a norma dell'articolo 35, paragrafo 3, del regolamento (UE) 2022/2065*, al par. 63

Sono emblematici, a livello internazionale, alcuni casi in cui è emerso con forza il tema dell'uso strategico e offensivo della disinformazione come strumento politico e antidemocratico, al punto da introdurre nel dibattito pubblico l'espressione di *first social media war*⁹. Tra gli episodi più significativi si segnalano il caso *Pizzagate* e la teoria cospirativa *QAnon*¹⁰, le dinamiche legate alla piattaforma X e il ruolo assunto da Elon Musk nelle recenti consultazioni elettorali statunitensi¹¹, il caso mediatico di *Lisa F.*¹², nonché il noto scandalo internazionale che ha coinvolto *Cambridge Analytica*¹³.

Tali vicende esemplificano la portata eversiva e destabilizzante della disinformazione digitale rispetto alla tutela della democrazia e all'integrità dei processi elettorali. La sistematica diffusione di *deepfake*, specie quando orchestrata da attori statali esteri, determina un grave inquinamento del dibattito pubblico e compromette il pieno esercizio della libertà di espressione e del diritto all'informazione, essenziale per il corretto funzionamento di ogni ordinamento democratico¹⁴. Ne è un esempio la recente decisione della Corte costituzionale rumena¹⁵ di annullare le elezioni presidenziali per sospette interferenze russe legate a una campagna elettorale condotta dal candidato Georgescu, prevalentemente *online* e, precisamente, su *TikTok*¹⁶.

nel concetto di minaccia ibrida, vengono inclusi la disinformazione, la manipolazione delle informazioni, le ingerenze da parte di attori stranieri, ma anche gli attacchi informatici.

⁹ P. SUCIU, *Is Russia's Invasion of Ukraine the First Social Media War*, in *Forbes*, March 2022.

¹⁰ Nel 2016, il caso *Pizzagate* nacque da interpretazioni distorte di email pubblicate da WikiLeaks, trasformate *online* in una falsa accusa secondo cui una pizzeria di Washington sarebbe stata la sede di un traffico di minori coinvolgente figure politiche democratiche; la teoria portò persino un uomo armato a irrompere nel locale senza trovare nulla. La teoria *QAnon* dal 2017 ampliò questo filone, sostenendo che un'élite globale di pedofili e satanisti controllerebbe la politica e i media, mentre un presunto informatore chiamato "Q" rivelerebbe *online* indizi criptici su un imminente "grande risveglio"; nessuna prova ha mai confermato tali affermazioni, ma il movimento ha influenzato gruppi estremisti e dinamiche politiche reali.

¹¹ Elon Musk ha assunto un ruolo attivo nelle recenti elezioni statunitensi, sostenendo politicamente Donald Trump e usando la piattaforma X per influenzare il dibattito pubblico. Ha finanziato un super-PAC che ha promosso iniziative pro-Trump e campagne mediatiche ad alto impatto. Ha promosso un'iniziativa definita 'lotteria', che offriva fino a un milione di dollari al giorno agli elettori registrati negli *swing States*, presentandola come un progetto civico di partecipazione democratica. Le autorità hanno però segnalato possibili violazioni della legge sul voto, interpretandola come un incentivo illegale agli elettori.

¹² Nel 2016, il caso di *Lisa F.* in Germania ha riguardato una ragazza russo-tedesca che denunciò un rapimento e stupro da parte di migranti. L'indagine dimostrò che la storia era inventata, ma la notizia iniziale generò scalpore mediatico, proteste e tensioni diplomatiche tra Germania e Russia.

¹³ Il noto scandalo *Cambridge Analytica* esplose nel 2018 quando emerse che la società aveva raccolto impropriamente i dati di milioni di utenti Facebook tramite un'app di quiz e che questi dati furono poi rivenduti e utilizzati per creare profili psicologici e influenzare campagne politiche, tra cui Brexit e le elezioni USA 2016.

¹⁴ Nella celebre sentenza *Handyside c. Regno Unito*, ric. 5493/72, sentenza del 7 dicembre 1976, para. 49, la Corte EDU ha posto in luce la duplice valenza della libertà di espressione, quale presupposto essenziale per lo sviluppo individuale, ma anche per il progresso della società nella sua collettività.

¹⁵ Romanian Constitutional Court, Judgement of 6 December 2024, no. 32. Si vedano D. VAIRA, *Trick or T(h)reat: disinformazione online e minacce ibride nel panorama europeo. Alcune considerazioni alla luce dell'annullamento delle elezioni in Romania*, in *SIDIBlog*, 29 dicembre 2024; R. VIORESCU, D. VARELA, *Constitutional Analysis of the Judgment of the Constitutional Court of Romania No. 32/2024 Annuling the Presidential Elections*, in *European Journal of Law and Public Administration*, 2024, pp. 243-260; L. DI ANSELMO, *La disinformazione online e i rischi per la democrazia: qualche considerazione sul ruolo del Digital Services Act alla luce delle elezioni presidenziali in Romania*, in *Ordine internazionale e diritti umani*, 3/2025.

¹⁶ Il 17 dicembre 2024 la Commissione europea ha avviato un procedimento formale nei confronti di *TikTok* per verificare se, nel contesto delle elezioni rumene, la piattaforma abbia rispettato gli obblighi ad essa incombenti ai sensi del DSA in materia di mitigazione dei rischi sistemici riguardanti il regolare svolgimento del

Nei casi più gravi, tali condotte possono arrivare ad integrare una violazione del principio di non ingerenza internazionale¹⁷.

Non sorprende, dunque, che la strategia europea di contrasto alla disinformazione si sia concentrata principalmente sul contenimento dell'impatto di influenze di matrice esterna, particolarmente in ambiente digitale.

È invero da rilevare che, in un simile scenario, pratiche consolidate nell'ambito *offline* – come il *gerrymandering* politico¹⁸, ossia la manipolazione strategica dei confini elettorali tramite tecniche di *packing* e *cracking* – assumono, *mutatis mutandis*, una nuova e più insidiosa dimensione nell'ecosistema digitale. Ciò in quanto, se nel contesto tradizionale il *gerrymandering* si realizza attraverso la ridefinizione territoriale di distretti elettorali fisici in un numero determinato e limitato, in ambito digitale la segmentazione degli elettori avviene attraverso frontiere non visibili, modellate dagli algoritmi sulla base delle interazioni e dei dati personali raccolti *online*, con una portata molto più estesa e pervasiva. Tale redistribuzione virtuale della popolazione, che può essere opportunamente descritta come *gerrymandering* digitale, da intendersi come la manipolazione degli spazi informativi e dei flussi algoritmici per influenzare selettivamente la visibilità politica degli utenti, consente pratiche avanzate di *microtargeting* e *retargeting* politico, con la conseguente esposizione mirata di gruppi specifici a determinati messaggi e la possibile esclusione di altri segmenti dall'arena informativa e partecipativa. Ne deriva un rischio di alterazione sistematica dell'equità della competizione democratica, che può realizzarsi attraverso l'azione poco trasparente di piattaforme e algoritmi, capaci di replicare, e talvolta accentuare, le distorsioni già note del sistema rappresentativo¹⁹. In un simile contesto, ciò che può emergere è una nuova forma di manipolazione del consenso – difficilmente regolabile e spesso opaca – che può incidere negativamente sull'imparzialità dei processi elettorali. Tale imparzialità rappresenta una precondizione imprescindibile per la tutela dell'ordine democratico, nel quale le regole e i flussi informativi che sostengono la competizione politica non devono garantire vantaggi sistematici ad alcun soggetto. Solo così è possibile assicurare pari opportunità di partecipazione, la libera formazione delle preferenze e decisioni collettive che rispecchino realmente la volontà pubblica, anziché gli interessi di posizioni di potere consolidate. In assenza di tali garanzie, si rischia di alimentare divisioni nell'elettorato e di alterare l'uguaglianza sostanziale nell'accesso alle informazioni.

Trovandosi di fronte a sfide tanto complesse, l'Unione europea, nel tentativo di contrastare il fenomeno della disinformazione, ha adottato strategie progressivamente più articolate²⁰. A partire dalla prima fase, avviata nel 2015 con lo scoppio del conflitto russo-

processo elettorale. Si vedano J. ALBERT, *TikTok and the Romanian elections: A stress test for DSA enforcement*, in *DSA Observatory*, 20 December 2024; J. BARATA, E. LAZĂR, *Will the DSA Save Democracy? The Test of the Recent Presidential Election in Romania*, in *Tech Policy Press*, 27 January 2025; A. IANNOTTI DELLA VALLE, *Libertà di espressione e valori democratici alla prova dei social media: il DSA e un nuovo caso TikTok europeo*, in *federalismi.it*, n. 13/2025, pp. 84-101.

¹⁷ La stessa Commissione nella Comunicazione sul *Piano d'azione per la democrazia europea*, cit., ha riconosciuto che la crescente digitalizzazione delle campagne elettorali e l'uso delle piattaforme *online* hanno esposto i processi democratici a nuove forme di manipolazione, disinformazione e vulnerabilità, tra cui finanziamenti opachi, l'*hacking* o il *defacing* di siti *web*, attacchi informatici alle infrastrutture elettorali e diffusione di messaggi polarizzanti tramite algoritmi poco trasparenti.

¹⁸ Il termine deriva dal governatore E. Gerry, che all'inizio del XIX secolo, ridisegnò i confini elettorali del Massachusetts in modo irregolare per favorire la propria rielezione, includendo le aree a lui favorevoli ed escludendo quelle ostili.

¹⁹ E. JORDAN, *How Computers turned Gerrymandering Into a Science*, in *New York Times*, 6 ottobre 2017.

²⁰ Si veda, O. POLLICINO, P. DUNN, *Intelligenza artificiale e democrazia*, cit., p. 129 ss.

ucraino, incentrata sulla risposta alla propaganda filorussa tramite la *East StratCom Task Force*²¹, si è passati a una stagione di *soft law* inaugurata dalla comunicazione della Commissione dell'aprile 2018²², che ha coinvolto le piattaforme *online* con il Codice di buone pratiche²³ e il sostegno all'autoregolamentazione. Constatata però la frammentazione normativa a livello nazionale²⁴ e l'inefficacia di tali iniziative, si è infine giunti a prediligere un approccio regolatorio più incisivo e armonizzato, segnato dall'adozione di strumenti come il *Digital Services Act* (DSA), il *Political Advertising Regulation* (PAR) e l'*AI Act*, che si esamineranno nel prosieguo, o di ulteriori atti fondamentali come il *Media Freedom Act* e il *Data Act*, su cui però non è possibile soffermarsi in questa sede.

3. Il quadro giuridico europeo di riferimento: DSA, PAR e AI Act alla luce della Carta dei diritti fondamentali

L'Unione europea, pur nel rispetto delle identità nazionali degli Stati membri²⁵, annovera il principio democratico tra i propri valori fondamentali²⁶, riconoscendolo quale criterio guida dell'azione di governo europea²⁷. Con l'entrata in vigore del Trattato di Lisbona, il principio democratico trova una declinazione rafforzata nella complementarità tra democrazia rappresentativa e partecipativa, come sancito agli articoli 10 e 11 TUE. Questa impostazione riflette la consapevolezza che la democrazia rappresentativa, pur rimanendo imprescindibile, non sia da sola sufficiente a garantire la piena attuazione del fondamento

²¹ La East StratCom Task Force (ESCTF) è un'unità istituita all'interno del Servizio europeo per l'azione esterna, incaricata di sviluppare strategie di comunicazione volte a rendere più efficace la diffusione e la promozione delle politiche e dei valori dell'Unione Europea nei Paesi dell'Europa orientale – comprendendo anche Armenia, Azerbaigian, Bielorussia, Georgia, Moldova e Ucraina – e, più in generale, nell'area che include anche la Russia. Dal 2022, il suo mandato si è ampliato fino a includere attività di contrasto e verifica delle campagne di disinformazione provenienti dalla Cina. Il progetto più importante di questa task force è *EUvsDisinfo*, un sito *web*, in tredici lingue, che include un database di articoli e media che ESCTF ha valutato che forniscano informazioni false, distorte o parziali nel loro contenuto, reperibile *online*.

²² Comunicazione della Commissione, COM/2018/236 final del 26 aprile 2018, *Contrastare la disinformazione online: un approccio europeo*.

²³ Codice di buone pratiche dell'UE sulla disinformazione, 20 settembre 2018. Si veda sul punto M. MONTI, *The EU Code of Practice on Disinformation and the Risk of the Privatisation of Censorship*, in S. GIUSTI, E. PIRAS (a cura di), *Democracy and Fake News: Information Manipulation and Post-Truth Politics*, Londra, 2020, pp. 214-225; G. MORGESE, *Il contrasto alla disinformazione originata da ingerenze straniere nell'Unione europea*, in M. Messina (a cura di), *Cittadinanza e stato di diritto per un'unione europea più forte*, Napoli, 2024, p. 89 ss.

²⁴ Da menzionare il *Netzwerkdurchsetzungsgesetz, Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken* volto a contrastare la diffusione di contenuti illeciti in rete si veda V. CLAUSSEN, *Fighting hate speech and fake news. The Network Enforcement Act (NetzDG) in Germany in the context of European Legislation*, in *Rivista del diritto dei media*, 3, pp. 110-136.

²⁵ Art. 4, par. 2 TUE.

²⁶ Art. 2 TUE. Si vedano A. PIZZORUSSO, *Il patrimonio costituzionale europeo*, Bologna 2002; U. VILLANI, *Istituzioni di diritto dell'Unione europea*, Bari, 2024, 7° ed., p. 43 ss. La Corte di giustizia, nelle note sentenze gemelle in materia di tutela dello Stato di diritto, causa C-156/21, *Ungheria c. Parlamento europeo e Consiglio*, sentenza del 16 febbraio 2022; causa C-157/21, *Polonia c. Parlamento europeo e Consiglio*, sentenza del 16 febbraio 2022, ha definito questi valori come espressione dell'essenza stessa dell'identità costituzionale europea.

²⁷ Sul punto si veda G. ZACCARONI, *Intelligenza artificiale e principio democratico: riflessioni a margine dell'emersione di un quadro normativo europeo*, in *Quaderni AISDUE*, 2/2024.

democratico dell'integrazione europea, rendendo necessario il rafforzamento degli strumenti di partecipazione diretta dei cittadini al processo decisionale dell'Unione²⁸.

Eppure, gli sforzi compiuti a livello europeo rischiano di essere vanificati dall'emergere di nuove tecnologie e dalle pratiche, talvolta esplicite e talora meno visibili, riconducibili a quella forma di *gerrymandering* digitale richiamata in precedenza, che finisce per attribuire un potere sproporzionato a chi dispone delle risorse per impiegare sofisticati strumenti di *targeting* o sfruttare i *bias* inevitabilmente presenti nei sistemi.

Un dato appare, infatti, incontrovertibile, il panorama digitale rimane saldamente nelle mani delle *Tech Giants*, quasi sempre con sede in Stati terzi, che hanno consolidato prassi di autoregolamentazione e imposto regole proprie, spesso al di fuori dei circuiti democratici²⁹ e in costante competizione con i poteri pubblici. Questo scenario pone un duplice rischio: da un lato, le piattaforme possono esercitare un'influenza determinante sulle scelte collettive; dall'altro, possono adottare forme di censura discrezionale sui contenuti, con effetti potenzialmente lesivi della libertà di espressione e delle dinamiche democratiche³⁰. In tale contesto, il principio di *net neutrality* – originariamente di natura tecnica – assume una rilevanza giuridica crescente, considerando che le regole della rete vengono dettate unilateralmente dalle piattaforme, che però per lungo tempo, non sono state ritenute responsabili dei contenuti immessi in rete³¹.

In un simile quadro eterogeneo, le politiche digitali europee si sono orientate nella direzione di un approccio regolatorio rinnovato³² volto a superare la tradizionale logica di intervento verticale e settoriale a favore di una regolamentazione più orizzontale e trasversale³³.

Il primo strumento che, nella prospettiva della presente analisi, appare opportuno esaminare è il DSA³⁴. L'obiettivo del regolamento è quello di assicurare un ambiente digitale sicuro e affidabile, in cui risultino garantiti i diritti riconosciuti dalla Carta dei diritti fondamentali³⁵ attraverso una disciplina articolata della moderazione dei contenuti e l'introduzione di un complesso di regole e procedure volte ad assicurare il rispetto dei principi

²⁸ Sul punto si veda G. MORGESE, *Principio e strumenti della democrazia partecipativa nell'Unione europea*, in E. Triggiani (a cura di), *Le nuove frontiere della cittadinanza europea*, Bari, 2011, p. 37 ss.

²⁹ Cfr. M. MONTI, *Privatizzazione della censura e Internet Platforms: la libertà d'espressione e i nuovi censori dell'agorà digitale*, in *Rivista italiana di informatica e diritto*, 1/2019, p. 37 ss.; M. BETZU, *I baroni del digitale*, Napoli, 2022, p. 34 ss. Si veda altresì la pronuncia della US Supreme Court, *Biden v. Knight First Amend. Inst. at Colum. Univ.*, No. 20-197, 5 April 2021, (J. Thomas, concurring opinion), 141 S. Ct. at 1222.

³⁰ Si vedano, *ex multis*, M. MANETTI, *Regolare Internet*, in *MediaLaws*, 2/2020, p. 36 ss.; M. BETZU, *Poteri pubblici e poteri privati nel mondo digitale*, in *Rivista del Gruppo di Pisa*, 2/2021, p. 172 ss.

³¹ Si veda M. C. GIRARDI, *Libertà e limiti della comunicazione nello spazio pubblico digitale*, in *federalismi.it*, 2024.

³² Per una ricognizione del quadro giuridico generale si veda G. CAGGIANO, *Il quadro normativo del mercato unico digitale*, in *Eurojus*, fascicolo speciale Mercato Unico Digitale, dati personali e diritti fondamentali, (a cura di) F. Rossi Dal Pozzo, 2020, p. 13 ss.

³³ O. POLLICINO, P. DUNN, *Intelligenza artificiale e democrazia*, cit., p. 129 ss.

³⁴ Regolamento (UE) 2022/2065 del 19 ottobre 2022, relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali). Sul tema disinformazione e DSA si vedano, *ex multis*, E. BIRRITTERI, *Contrasto alla disinformazione, Digital Services Act e attività di private enforcement: fondamento, contenuti e limiti degli obblighi di compliance e dei poteri di autonormazione degli operatori*, in *MediaLaws*, 2/2023, p. 73; L. D'AGOSTINO, *Disinformazione e obblighi di compliance degli operatori del mercato digitale alla luce del nuovo Digital Services Act*, in *MediaLaws*, 2/2023, p. 33 ss.; N. ZINGALES, *The DSA as a Paradigm Shift for Online Intermediaries' Due Diligence*, in J. VAN HOBOK, J.P. QUINTAIS, N. APPELMAN, R. FAHY, I. BURI, M. STRAUB (eds.), *Putting the DSA into Practice. Enforcement, Access to Justice and Global Implications*, Berlin, 2023, p. 211 ss.

³⁵ Sul tema della salvaguardia dei diritti fondamentali e del loro bilanciamento si veda estensivamente F. FERRI, *Il bilanciamento dei diritti fondamentali nel mercato unico digitale*, Torino, 2022.

di legalità, trasparenza e limitazione della discrezionalità dei fornitori di servizi *online*³⁶. Attraverso specifiche misure di filtraggio, linee guida e protocolli, il DSA mira a creare un quadro normativo europeo uniforme, distinguendo chiaramente le responsabilità dei fornitori di servizi. Il regolamento introduce severi obblighi di trasparenza e una definizione di “moderazione dei contenuti”³⁷, imponendo alle piattaforme di pubblicare dettagli sui criteri e parametri utilizzati, pur mantenendo il regime di esonero dalla responsabilità diretta dei *provider*. L’ambito di applicazione si estende a tutti i servizi di intermediazione destinati a utenti nell’Unione europea, indipendentemente dalla sede del fornitore, purché esista un rapporto significativo con i destinatari³⁸. Inoltre, i fornitori sono tenuti a rispettare obblighi di vigilanza graduati in funzione del servizio e delle dimensioni dell’intermediario. Nello specifico, VLOPs e VLOSEs sono soggetti sia agli obblighi generali, sia a requisiti aggiuntivi, tra cui la pubblicazione delle *policy* e delle procedure di moderazione – algoritmica e umana –, nonché la predisposizione di *report* periodici sulle attività svolte. Il DSA accresce altresì le tutele per gli utenti, consentendo di segnalare contenuti illeciti e imponendo alle piattaforme trasparenza nelle raccomandazioni e nella pubblicità, e vietando pratiche ingannevoli nella gestione delle interfacce³⁹. Mira, inoltre, a garantire la piena consapevolezza circa le opzioni a disposizione nell’utilizzo dei servizi *online*, anche mediante procedure di *disclaimer*⁴⁰ e richiede alle piattaforme di rendere noti i parametri utilizzati nei sistemi di raccomandazione e di indicare chiaramente quando un contenuto costituisce pubblicità, specificando *sponsor* e committenti⁴¹.

Permangono, tuttavia, alcune criticità nell’impianto sanzionatorio delineato dal DSA che si configura come un sistema multilivello fondato sulla combinazione tra competenze nazionali e poteri centralizzati in capo alla Commissione. Agli Stati membri è attribuito il compito di definire e applicare le sanzioni nei confronti della generalità dei fornitori di servizi intermediari, elaborando regimi nazionali che garantiscano effettività, proporzionalità e dissuasività⁴². La Commissione, invece, assume un ruolo diretto e centrale nei confronti delle piattaforme *online* e dei motori di ricerca di dimensioni molto grandi, sui quali esercita poteri investigativi autonomi e può irrogare sanzioni pecuniarie di particolare rilevanza. Proprio questa ripartizione di competenze tra livello europeo e nazionale genera però il rischio di applicazioni disomogenee, alimentando disparità tra ordinamenti e asimmetrie tra operatori “sistemici” e non, con possibili effetti negativi sulla coerenza del sistema e sulla tutela uniforme degli utenti.

Come precedentemente evidenziato, un ulteriore elemento di primaria importanza per la salvaguardia dell’assetto democratico concerne la disciplina della pubblicità politica, ambito nel quale si assiste a un crescente impiego di sofisticate tecniche di profilazione degli utenti.

³⁶ Si vedano O. POLLICINO, G. DE GREGORIO, M. BASSINI, *Verso il Digital Services Act: problemi e prospettive. Presentazione del simposio*, in *MediaLaws*, 23 novembre 2020; G. CAGGIANO, *Il contrasto alla disinformazione tra nuovi obblighi delle piattaforme online e tutela dei diritti fondamentali nel quadro del Digital Service Act e della co-regolamentazione*, in *Papers di diritto europeo*, 2021, pp. 45-72 e G. CAGGIANO, G. CONTALDI, P. MANZINI (a cura di), *Verso una legislazione europea sui mercati digitali*, Bari, 2021.

³⁷ Art. 3, par. 1, lett. t).

³⁸ Si veda F. MAROGIU BUONAIUTI, *L’ambito di applicazione territoriale del Digital Markets Act e del Digital Services Act*, in G. CAGGIANO, G. CONTALDI, P. MANZINI (a cura di), *Verso una legislazione europea sui mercati digitali*, cit., p. 171 ss.

³⁹ Art. 25 DSA.

⁴⁰ Art. 15, par. 1, lett. e) DSA.

⁴¹ Art. 26 DSA. Si veda M.R. ALLEGRI, *Il futuro digitale dell’Unione europea: nuove categorie di intermediari digitali, nuove forme di responsabilità*, in *Rivista italiana di informatica e diritto*, 2/2021, p. 14 ss.

⁴² Art. 52 DSA.

Tali strumenti consentono l'invio di messaggi personalizzati e mirati, finalizzati a massimizzare l'efficacia persuasiva della comunicazione politica e a incidere in modo più determinante sulla formazione del consenso elettorale⁴³.

L'Unione europea ha alzato il livello di attenzione sul tema della disinformazione dopo i casi delle elezioni USA 2016 e della *Brexit*, raccomandando l'adozione di misure specifiche⁴⁴. Tale impegno si è concretizzato nell'adozione del regolamento sulla trasparenza e il *targeting* della pubblicità politica (PAR)⁴⁵, adottato sulla base degli articoli 16 e 114 TFUE. Il PAR introduce standard uniformi in materia di trasparenza, obblighi di *due diligence* e gestione dei dati per la pubblicità politica, regolamentando in particolare l'impiego di tecniche di *targeting* fondate sul trattamento di dati personali. Prevede inoltre l'istituzione di un archivio pubblico accessibile, volto a garantire la tracciabilità e la pubblicità delle inserzioni politiche *online* diffuse nell'Unione europea⁴⁶.

All'interno del regolamento viene proposta una definizione di "pubblicità politica" particolarmente estesa, che si sviluppa su due livelli: soggettivo, includendo non solo partiti, candidati e attori istituzionali, ma anche *sponsor*, editori e piattaforme coinvolte; ed oggettivo, considerando come pubblicità politica qualsiasi messaggio, a prescindere dal soggetto promotore, che possa influenzare un'elezione, un *referendum* o un processo legislativo a qualsiasi livello⁴⁷. Ne consegue che una comunicazione può assumere la natura di pubblicità politica sia in ragione dello *sponsor*, sia in considerazione degli effetti perseguiti, come accade, ad esempio, per le cosiddette *issue-based ads* o *social issue ads*. Questa ampiezza definitoria comporta il rischio di ricomprendere anche forme di comunicazione non strettamente politiche, come l'attivismo sociale, il giornalismo o le semplici espressioni civiche, sollevando così dubbi sulla possibilità di restrizioni eccessive alla libertà di espressione.

Al centro del PAR vi sono gli obblighi di identificazione e trasparenza delle inserzioni politiche⁴⁸. Come visto nel caso del DSA⁴⁹, anche nel PAR la trasparenza assume un rilievo fondamentale⁵⁰ per rendere la pubblicità politica più facilmente riconoscibile, con una probabile riduzione dei fenomeni di disinformazione e altre manipolazioni. A motivo di ciò, ogni pubblicità politica deve recare un'etichetta identificativa, accompagnata da una nota di trasparenza che indichi i dettagli relativi allo *sponsor* dell'annuncio e al procedimento elettorale di riferimento. Viene richiesta la comunicazione di informazioni sia al pubblico in generale, sia alle autorità di regolamentazione, sia agli utenti destinatari delle singole inserzioni e agli altri soggetti interessati⁵¹. Le piattaforme di grandi dimensioni sono soggette a obblighi aggiuntivi di segnalazione e gestione delle notifiche riguardanti inserzioni non conformi⁵².

⁴³ Particolarmente significative sono pratiche promosse negli Stati Uniti come il *voter isolation* e la commercializzazione di database elettorali da parte di società come Aristotle International - che ha venduto dati su 150 milioni di elettori per scopi di marketing e propaganda politica - hanno sollevato ampie discussioni sul rischio di sfruttamento e manipolazione dei dati personali in ambito elettorale.

⁴⁴ Si veda per approfondimenti M. Z. VAN DRUNEN, N. HELBERGER, R. Ó FATHAIGH, *The Beginning of EU Political Advertising Law: Unifying Democratic Visions through the Internal Market*, in *International Journal of Law and Information Technology*, 2022, p. 181 ss.

⁴⁵ Regolamento (UE) 2024/900 del 13 marzo 2024 relativo alla trasparenza e al *targeting* della pubblicità politica.

⁴⁶ Art. 1 par. 1 PAR.

⁴⁷ Art. 3 par. 2 e considerando 22 e 23 PAR.

⁴⁸ Artt. 7 e 15 e si vedano altresì i considerando 44 e 54 PAR.

⁴⁹ Articoli 26 e 39 DSA.

⁵⁰ Art. 6, par. 1 PAR.

⁵¹ Art. 11, par. 1 e art. 12 PAR.

⁵² Si veda S. ESKENS, *The role of the Political Advertising Regulation and European Media Freedom Act in the EU's anti-disinformation approach*, *Working Paper*, 31 August 2024, reperibile online.

Si aggiunga che, opportunamente, il PAR limita l'uso delle strategie di *targeting* e *ad-delivery*⁵³. Tali pratiche sono consentite solo previo consenso esplicito dell'interessato⁵⁴, esclusa la profilazione su dati sensibili, come prescritto anche nel DSA. Rimane tuttavia incerto se il PAR consenta forme di *targeting* che, pur non prevedendo la profilazione, si basino comunque su dati sensibili. Queste restrizioni intendono prevenire la manipolazione e le interferenze tramite l'uso delle vulnerabilità dei singoli soggetti, posto che il trattamento dei dati personali a fini di pubblicità politica può incidere negativamente sui diritti e sulle libertà individuali, oltre a rappresentare un potenziale rischio per il corretto funzionamento del processo democratico.

I titolari del trattamento sono, inoltre, tenuti a garantire il massimo livello di trasparenza nell'impiego di algoritmi e sistemi di IA utilizzati per il *targeting* pubblicitario, adottando *policy* chiare e informando gli utenti circa i criteri adottati. Tale obbligo si pone, ancora una volta, in continuità con quanto previsto da DSA⁵⁵ e, in questo caso, anche dal GDPR⁵⁶, che impongono di informare gli interessati in merito alle logiche sottese alle decisioni automatizzate e ai principali parametri impiegati nei sistemi di raccomandazione e di pubblicità.

Si prevede che la Commissione istituisca un archivio pubblico delle inserzioni politiche *online* diffuse nell'Unione, diretti a cittadini europei o residenti⁵⁷. Inoltre, al fine di adempiere agli obblighi previsti dal PAR, i fornitori di pubblicità politica che, ai sensi del DSA, sono considerati VLOPs o VLOSEs, dovranno garantire la disponibilità e la trasparenza di tali inserzioni sia tramite tale archivio europeo sia attraverso propri registri esistenti⁵⁸.

Benché il quadro normativo risulti articolato e meritevole di apprezzamento, permangono nondimeno alcune criticità. Sebbene, infatti, il PAR persegua la trasparenza delle inserzioni politiche, affida la responsabilità primaria di identificazione allo *sponsor*, imponendo alle piattaforme digitali requisiti solo minimi in materia di controllo. Questa impostazione potrebbe consentire agli inserzionisti che operano in maniera non trasparente di eludere facilmente gli obblighi di legge, omettendo semplicemente la dichiarazione della natura politica dell'annuncio e aggirando così i meccanismi di trasparenza e rendicontazione previsti. L'assenza di incisivi obblighi di *due diligence* per le piattaforme espone il sistema al rischio di manipolazione informativa, soprattutto nei casi di utilizzo illecito dei dati personali o di discrepanza tra fondi impiegati e dichiarati. In tale contesto, il fenomeno dei 'falsi negativi' – ovvero annunci politici non riconosciuti né dalla piattaforma, né dallo *sponsor* – sottolinea la necessità di rafforzare gli obblighi di identificazione su entrambi i fronti.

Alla mitigazione di una simile criticità potrebbe in parte soccorrere il DSA che, sotto il profilo della trasparenza, offre un quadro di tutela più integrato, imponendo alle VLOPs e ai VLOSEs obblighi specifici di valutazione e mitigazione dei rischi sistemici, inclusa la mancata identificazione delle inserzioni politiche. In questo senso, particolarmente rilevanti sono le disposizioni che qualificano come rischi sistemici i danni al dibattito pubblico e alle procedure elettorali, imponendo misure correttive sulle modalità di diffusione degli annunci⁵⁹. Nella stessa prospettiva, appare centrale anche l'obbligo per le grandi piattaforme

⁵³ Considerando 77 PAR.

⁵⁴ Art. 4 GDPR.

⁵⁵ Art. 39 par. 1, lett. f) DSA.

⁵⁶ Art. 22 GDPR.

⁵⁷ Art. 13, par. 1, PAR.

⁵⁸ Art. 13, par. 2 e art. 16 PAR.

⁵⁹ Articoli 34 e 35 DSA.

di istituire e pubblicare un registro degli annunci pubblicitari, comprensivo di informazioni sui criteri di *targeting* e sulla profilazione degli utenti. Questo sistema di trasparenza rafforzata può contribuire a rendere più chiari anche i meccanismi sottesi alla pubblicità politica, limitando il rischio di manipolazione e garantendo una maggiore *accountability* degli intermediari digitali.

Va rilevato che la complessità dell'ecosistema digitale è ulteriormente accresciuta dal fatto che il funzionamento delle piattaforme si fonda in larga misura su sistemi algoritmici. Questo rende imprescindibile, nell'ottica della presente disamina, anche alcune considerazioni in merito alla disciplina contenuta all'interno del regolamento sull'intelligenza artificiale (*AI Act*). Tali sistemi, infatti, si prestano a molteplici usi potenzialmente lesivi della qualità del dibattito pubblico, favorendo la produzione automatizzata di contenuti disinformativi e, in alcuni casi, anche di *output* frutto di cosiddette 'allucinazioni', ossia informazioni meramente misinformative⁶⁰.

Si tratta di un atto notoriamente complesso, il cui esame di dettaglio esula dalla presente disamina, nondimeno per ciò che riguarda la tutela dell'assetto democratico l'*AI Act* si affida eminentemente a due strumenti, uno di natura trasversale e orizzontale, che si sostanzia nell'approccio basato sul rischio – attraverso il quale si vietano o si sottopongono a obblighi stringenti i sistemi con portata lesiva particolarmente alta – e l'altro, più specifico, ovvero la valutazione dell'impatto sui diritti fondamentali, sulla democrazia e lo Stato di diritto⁶¹. Il sistema di gestione del rischio delineato dall'*AI Act*, articolato nelle quattro fasi di identificazione e analisi, valutazione, monitoraggio successivo all'immissione sul mercato e adozione delle misure di mitigazione, si propone come risposta normativa alle minacce dell'IA per l'assetto democratico, rafforzandosi nel coordinamento con strumenti orizzontali come i già esaminati DSA e PAR. Pur sussistendo alcuni profili di non piena coerenza normativa, il quadro così delineato mira a bilanciare tutela dei diritti individuali e salvaguardia degli interessi pubblici fondamentali, tra cui la democrazia.

La centralità di questi valori emerge chiaramente già dalla parte introduttiva del regolamento: nei consideranda 1, 2 e 8 si sottolinea la necessità che l'intelligenza artificiale rispetti i valori europei; nei consideranda 61 e 62 si richiama l'attenzione sui rischi legati ai sistemi di IA ad alto rischio nei processi democratici e giudiziari, mentre il considerando 120 evidenzia l'importanza del coordinamento con il DSA, soprattutto in relazione ai rischi per il dibattito pubblico e i processi elettorali.

Sul piano operativo, il sistema di gestione del rischio per i sistemi di IA ad alto rischio, disciplinato dall'art. 9, deve accompagnare l'intero ciclo di vita dell'IA e si articola in quattro fasi fondamentali: identificazione e analisi dei rischi per la salute, la sicurezza e i diritti fondamentali; valutazione e stima dei rischi derivanti da un uso improprio; monitoraggio dei rischi successivi all'immissione sul mercato; adozione di misure appropriate per mitigarli. Si tratta di un processo dinamico e ininterrotto che impone aggiornamenti e revisioni costanti, e che riflette la volontà del legislatore europeo di garantire un elevato livello di salvaguardia, soprattutto nei confronti dei soggetti più deboli.

⁶⁰ Per allucinazione algoritmica si intende la produzione, da parte di un modello computazionale, di *output* non aderenti ai dati, di contenuti non verificati o non supportati, risultanti da errori di generalizzazione, insufficienze del *training set* utilizzati o limiti nei meccanismi di inferenza. Si veda la Communication to the Commission, 26.3.2024 C(2024) 2121 final, *Approval of the content of a draft Communication from the Commission on Guidelines for providers of Very Large Online Platforms and Very Large Online Search Engines on the mitigation of systemic risks for electoral processes pursuant to the Digital Services Act (Regulation (EU) 2022/2065)*.

⁶¹ G. ZACCARONI, *Intelligenza artificiale e principio democratico*, cit., p. 23 ss.

Un ulteriore elemento di rafforzamento della tutela è rappresentato dall'art. 27, il quale impone agli enti pubblici, agli enti privati che offrono servizi di interesse pubblico e ai *deployer* di sistemi di IA ad alto rischio in determinati settori, l'obbligo di effettuare una valutazione d'impatto sui diritti fondamentali anteriormente al primo utilizzo del sistema. Tale valutazione, concepita per identificare e prevenire potenziali impatti negativi sull'esercizio dei diritti garantiti dalla Carta dei diritti fondamentali dell'Unione europea, inclusi il principio democratico e lo Stato di diritto, deve descrivere le misure di mitigazione e sorveglianza adottate e prevedere azioni correttive in caso di danni. L'obbligo si applica soprattutto ai soggetti che utilizzano l'IA in ambiti sensibili e la valutazione va notificata all'autorità di vigilanza, potendo essere integrata con le valutazioni d'impatto *privacy* già previste dalla normativa europea.

In linea con questi obiettivi, l'art. 74, par. 8, stabilisce che, nei settori particolarmente sensibili per la vita democratica – come l'attività di contrasto, la gestione delle frontiere, la giustizia e la stessa democrazia – l'utilizzo di sistemi di intelligenza artificiale ad alto rischio sia soggetto a una vigilanza rafforzata da parte delle autorità nazionali per la protezione dei dati personali o di altri organismi dotati delle necessarie garanzie di indipendenza. Al tempo stesso, si tutela esplicitamente l'autonomia del potere giudiziario, quale presidio imprescindibile dello Stato di diritto, escludendo espressamente qualsiasi forma di interferenza. Inoltre, l'art. 77 attribuisce alle autorità pubbliche nazionali incaricate della tutela dei diritti fondamentali il potere di accedere a tutta la documentazione sui sistemi di IA ad alto rischio, chiedere verifiche tecniche, ove necessario, e coordinarsi con le autorità di vigilanza del mercato, rafforzando così le garanzie di trasparenza, *accountability* e controllo democratico.

Nel complesso, dunque, l'approccio delineato dall'*AI Act* pone l'accento sulla necessità di garantire un bilanciamento tra la promozione dell'innovazione e la salvaguardia dell'assetto democratico. Tuttavia, tale impostazione non è esente da criticità, poiché richiede di temperare spinte regolatorie ed esigenze di sviluppo senza compromettere, da un lato, la competitività del settore e, dall'altro, la tutela dei diritti fondamentali e delle dinamiche istituzionali che sostengono il sistema democratico.

4. Conclusioni

Alla luce di quanto sopra, alcune riflessioni conclusive possono svolgersi in punto di coerenza normativa e di coordinamento degli atti esaminati. In assenza di procedure integrate e di meccanismi di raccordo efficaci, il quadro giuridico delineato rischia di lasciare spazi di tutela scoperti proprio su aspetti essenziali per la protezione della democrazia, come la prevenzione di *deepfake* elettorali e la manipolazione automatizzata del consenso. A ciò si sommano le incertezze sulla ripartizione delle competenze tra autorità nazionali, Commissione europea e nuove strutture di supervisione, con il rischio che l'azione di tutela risulti frammentata o, al contrario, eccessivamente burocratizzata e, dunque, poco tempestiva.

Sotto un'altra prospettiva, l'*AI Act* – così come il PAR – solleva il delicato problema del bilanciamento tra il contrasto alle forme più insidiose di manipolazione digitale e la salvaguardia della libertà di espressione e del pluralismo del dibattito pubblico, filtri automatizzati e restrizioni non adeguatamente calibrati rischiano, infatti, di comprimere

ingiustificatamente lo spazio del confronto democratico. Il DSA, pur fissando obblighi di trasparenza e gestione responsabile dei contenuti, risulta troppo generico sulla pubblicità politica, lasciando ampi margini di discrezionalità alle piattaforme. Il PAR colma solo parzialmente questo vuoto, imponendo requisiti di etichettatura e trasparenza, ma demandando quasi integralmente agli *sponsor* la qualificazione delle inserzioni come politiche, mentre i controlli a carico delle piattaforme restano minimi, rende agevole, per chi voglia eludere il sistema, aggirare le misure di trasparenza semplicemente omettendo di dichiarare la natura politica di un messaggio.

Non meno significativo, nella prospettiva qui adottata, è il fatto che l'*AI Act*, nonostante l'ambizione di presidiare i rischi che le tecnologie di intelligenza artificiale possono comportare per i processi democratici, soffre di una certa indeterminatezza nella definizione di cosa costituisca effettivamente un "rischio per la democrazia". Questa vaghezza si traduce, ancora una volta, in una eccessiva discrezionalità affidata sia alle autorità di vigilanza sia agli operatori di mercato, con il rischio di produrre applicazioni disomogenee a livello nazionale.

Va ricordato inoltre che gli stringenti obblighi di gestione e valutazione del rischio, nonché di monitoraggio continuo, imposti alle IA ad alto rischio, risultano particolarmente onerosi, soprattutto per operatori di piccole dimensioni. Ne consegue che soltanto i grandi operatori sono realisticamente in grado di sostenere i costi della *compliance*, con la potenziale conseguenza di una riduzione del pluralismo informativo e, indirettamente, di un indebolimento delle garanzie democratiche. Inoltre, la disciplina sul *risk management* introdotta dall'*AI Act* fatica a coordinarsi in modo sistematico con le previsioni del DSA e del PAR, in particolare rispetto alla trasparenza algoritmica e alla gestione delle segnalazioni di contenuti illeciti o manipolativi. Quanto al primo profilo, il DSA prevede la spiegazione delle logiche algoritmiche per la raccomandazione dei contenuti; il PAR esige etichettatura e note di trasparenza per ogni inserzione politica; tuttavia la responsabilità di verifica è suddivisa tra più attori e la mancanza di controlli effettivi lato piattaforma può lasciare ampi spazi di opacità. Analogamente, i sistemi di segnalazione risultano frammentati tra i tre regolamenti e la possibilità di un intervento tempestivo su contenuti manipolativi o pubblicità non dichiarate resta limitata, soprattutto per i piccoli *publisher* o per i casi in cui l'IA impiegata non sia espressamente "ad alto rischio".

Questa segmentazione rischia di avere un impatto negativo non solo sulla coerenza del sistema, ma soprattutto sulla tutela effettiva della democrazia e dei diritti fondamentali. Pratiche come il *microtargeting*, la profilazione spinta e l'utilizzo distorto degli algoritmi possono portare a una banalizzazione del dibattito pubblico, a una selezione degli interlocutori politici sempre più ristretta e alla progressiva esclusione delle minoranze dai processi comunicativi. Ciò potrebbe condurre ad un'erosione del pluralismo e del principio democratico, aggravata dal rischio che soggetti dotati di mezzi quasi illimitati e del controllo su piattaforme influenti possano polarizzare il discorso pubblico, manipolando l'opinione collettiva anche attraverso pratiche riconducibili al precedentemente delineato *gerrymandering* digitale.

Il contemperamento dei diversi interessi in gioco resta però fondamentale poiché, in senso opposto, il crescente ricorso a procedure automatizzate di moderazione, se non adeguatamente supervisionato, rischia di comprimere in modo irragionevole la libertà di espressione. Vi è, infatti, il pericolo che filtri algoritmici non distinguano correttamente tra manifestazioni di legittimo dissenso politico e contenuti realmente manipolativi. Ciò può

tradursi in forme di censura arbitraria o, al contrario, in un'insufficiente rilevazione di contenuti lesivi del dibattito democratico.

In prospettiva giuridica, si pone, dunque, in termini particolarmente stringenti il tema del bilanciamento e della proporzionalità, poiché – come affermato anche dalla Corte europea dei diritti dell'uomo⁶² – eventuali restrizioni all'espressione politica sono legittime solo se necessarie e giustificate dalla necessità di tutelare effettivamente il processo democratico. L'attuale assetto regolatorio rischia, invece, di introdurre limitazioni indifferenziate, idonee ad incidere anche su iniziative civiche e comunicazioni di interesse pubblico, compromettendo così il nucleo essenziale delle libertà costituzionali.

Appare quindi urgente, per il futuro, rafforzare il coordinamento all'interno del delineato quadro normativo vigente, sviluppando procedure integrate di controllo, interoperabilità dei sistemi di segnalazione e di trasparenza e meccanismi di *audit* congiunti tra autorità, piattaforme e *stakeholder*.

Solo un approccio realmente sinergico e multilivello potrà garantire una *governance* digitale in grado di affrontare le molteplici sfide, capace di tutelare la libertà di espressione, il pluralismo e la qualità della democrazia europea nell'era delle piattaforme e dell'intelligenza artificiale.

⁶² Secondo la Corte EDU, elezioni libere e libertà di espressione costituiscono il fondamento di ogni democrazia. I due diritti sono tra loro interconnessi e si rafforzano reciprocamente, cfr., *ex multis*, *Orlovskaya Iskra c. Russia*, ric. n. 42911/08, sentenza del 21 febbraio 2017, par. 110. Tuttavia, in determinate circostanze, può essere necessario imporre alcune restrizioni alla libertà di espressione per garantire un giusto equilibrio, cfr. *Bowman c. Regno Unito*, ric. n. 141/1996/760/961, sentenza del 19 febbraio 1998, par. 43.



LORENZO DI ANSELMO*

INTELLIGENZA ARTIFICIALE, LOTTA ALLA DISINFORMAZIONE ONLINE E TUTELA DELLA DEMOCRAZIA NELL'UNIONE EUROPEA ALLA LUCE DELL'AI ACT

SOMMARIO: 1. Considerazioni preliminari sul rapporto tra intelligenza artificiale e disinformazione. – 2. Brevi cenni sulla struttura e sulle finalità dell'*AI Act*. – 3. Le pratiche vietate ai sensi dell'art. 5: quale rilievo ai fini del contrasto alla disinformazione *online*? – 4. Il contrasto alla disinformazione nell'ambito dei sistemi di intelligenza artificiale ad alto rischio. – 5. Gli obblighi di trasparenza per i sistemi di IA a rischio minimo e le disposizioni specifiche in materia di *deep fake*. – 6. Conclusioni.

1. Considerazioni preliminari sul rapporto tra intelligenza artificiale e disinformazione

In un'epoca caratterizzata dalla c.d. infodemia, ossia dalla diffusione virale e spesso incontrollata delle informazioni, l'integrazione dei sistemi di intelligenza artificiale (IA) negli strumenti di comunicazione digitale appare destinata ad impattare profondamente sul dibattito pubblico. Anche laddove il ricorso a tali sistemi non celi un intento malevolo, il loro utilizzo può favorire la circolazione di notizie false o non attendibili, amplificando gli effetti dannosi della disinformazione¹. Ciò perché, nell'epoca plasmata dalle reti virtuali, l'efficacia e la visibilità di un contenuto non dipendono dalla sua veridicità, quanto piuttosto dalla sua capacità di generare interazioni. Tale dinamica, premiando le notizie più divisive, ancorché non verificate, poiché più funzionali ad intercettare l'interesse degli utenti, tende ad accelerare il flusso di contenuti disinformativi². Non a caso, i sistemi algoritmici che governano le piattaforme digitali sono progettati per raggiungere gli utenti più sensibili a specifici contenuti, con lo scopo di attirarne l'attenzione e massimizzarne il coinvolgimento. Sotto questo profilo, l'IA può essere impiegata per finalità di profilazione, permettendo, specialmente attraverso tecniche di apprendimento automatico, di elaborare i dati, le

* Borsista di ricerca *post-doc* in Diritto dell'Unione europea, Università di Roma La Sapienza.

¹ In generale sul rapporto tra intelligenza artificiale e disinformazione, cfr. N. BONTRIDDER, Y. POULLET, *The Role of Artificial Intelligence in Disinformation*, in *Data & Policy*, n. 3, 2021, p. 3 ss. Per una prospettiva più prettamente giuridica, cfr. O. POLLICINO, P. DUNN, *Intelligenza artificiale e democrazia. Opportunità e rischi di disinformazione e discriminazione*, Milano, 2024; A. RUFFO, *Il disordine informativo e l'Intelligenza Artificiale; tra insidie e possibili strumenti di contrasto*, in *MediaLaws*, 2025, p. 407 ss.

² S. VOSOUGHI, D. ROY, S. ARAL, *The Spread of True and False News Online*, in *Science*, 2018, p. 1146 ss.

abitudini e le preferenze degli utenti al fine di proporre loro contenuti personalizzati e, quindi, più attrattivi³. Oltre ad amplificare il fenomeno della disinformazione *online*, l'IA, attraverso i c.d. modelli generativi, può essere sfruttata per produrre intenzionalmente contenuti falsi o, comunque, manipolati in modo tale da alterare la percezione della realtà. Tali sistemi, infatti, permettono di creare testi e immagini altamente realistici, difficili da distinguere persino dagli utenti più accorti: emblematico, in tal senso, il caso dei *deep fake*, video in cui personaggi pubblici sembrano compiere azioni o pronunciare parole in realtà mai avvenute.

Benché tali fenomeni minaccino costantemente di inquinare il dibattito pubblico⁴, ci sembra irrealistico immaginare di passare in rassegna tutte le notizie che circolano nella rete al fine di censurare quelle ritenute false, non solo perché la velocità di diffusione delle informazioni *online* renderebbe vana, o quantomeno parziale, siffatta operazione, ma anche perché un simile scenario presupporrebbe l'individuazione di criteri rigorosi sulla base dei quali valutare la veridicità di un contenuto, con la conseguenza di comprimere la libertà di espressione degli utenti⁵, tutelata, nell'ordinamento dell'Unione, dall'art. 11 della Carta dei diritti fondamentali. In ragione di ciò, non sorprende che le iniziative intraprese al riguardo dall'Unione europea si siano concentrate sui contenuti disinformativi suscettibili di pregiudicare i valori fondamentali dell'Unione di cui all'art. 2 TUE, *in primis* i diritti umani e i principi democratici⁶, specialmente allorché la circolazione di tali contenuti sia riconducibile a campagne disinformative architettate deliberatamente da attori stranieri⁷.

In questi casi, infatti, i sistemi di intelligenza artificiale possono essere sfruttati per promuovere narrazioni faziose e non aderenti alla realtà, aumentandone esponenzialmente la visibilità nella sfera digitale. Se pianificate in corrispondenza di importanti appuntamenti elettorali, quando i cittadini diventano particolarmente interessati alla ricezione di messaggi politici, queste campagne possono contribuire a manipolare l'opinione pubblica, influenzando la formazione del consenso popolare e, quindi, alterando il regolare

³ Sul tema, cfr. F. LAGIOIA, G. SARTOR, *Profilazione e decisione algoritmica: dal mercato alla sfera pubblica*, in *federalismi.it*, n. 11, 2020, p. 85 ss.

⁴ Una preoccupazione in tal senso è espressa, non a caso, anche nella Dichiarazione europea sui diritti e i principi digitali per il decennio digitale, la quale, nel declinare la libertà di espressione secondo le peculiarità proprie dell'ambiente digitale, precisa che le piattaforme digitali, «visto il ruolo svolto dai loro servizi nel plasmare l'opinione pubblica e il dibattito pubblico [...] dovrebbero attenuare i rischi derivanti dal funzionamento e dall'uso dei loro servizi, anche in relazione alle campagne di disinformazione e cattiva informazione» (Dichiarazione comune di Parlamento europeo, Consiglio e Commissione europea, *Dichiarazione europea sui diritti e i principi digitali per il decennio digitale*, 23 gennaio 2023, par. 15). Per un commento generale sulla Dichiarazione, si rinvia a P. DE PASQUALE, *Verso una Carta dei diritti digitali (fondamentali) dell'Unione europea?*, in *Il Diritto dell'Unione europea*, 2022, p. 163 ss.

⁵ B. BOTERO ARCILA, R. GRIFFIN, *Social media platforms and challenges for democracy, rule of law and fundamental rights*, Study requested by the LIBE Committee, Brussels, 2023, p. 77, reperibile *online*.

⁶ In effetti, nell'accezione accolta dall'Unione europea (cui si farà quindi riferimento in questa sede), non è il contenuto falso o ingannevole di per sé a qualificare una notizia come disinformazione, quanto la sua predisposizione ad arrecare un "pregiudizio pubblico", dunque un danno a un interesse di carattere generale per la collettività. Rileva, al riguardo, la definizione contenuta in una comunicazione della Commissione del 2018, secondo cui per disinformazione debba intendersi «un'informazione rivelatasi falsa o fuorviante concepita, presentata e diffusa a scopo di lucro o per ingannare intenzionalmente il pubblico, e che può arrecare un pregiudizio pubblico» (Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, *Contrastare la disinformazione online: un approccio europeo*, 26 aprile 2018, COM(2018) 236 fin.).

⁷ G. MORGESE, *Il contrasto alla disinformazione originata da ingerenze straniere nell'Unione europea*, in M. MESSINA (a cura di), *Cittadinanza e stato di diritto per un'unione europea più forte*, Napoli, 2024, p. 89 ss.

svolgimento della competizione elettorale⁸. La stessa Commissione europea ha messo in guardia dal fatto che, nel contesto delle elezioni, «la creazione e l'uso diffuso di sistemi di intelligenza artificiale [...] potrebbero comportare specifici rischi sistemici»⁹. Ne è un esempio quanto accaduto a fine 2024 in Romania, dove la Corte costituzionale ha annullato il primo turno delle elezioni presidenziali sostenendo che le attività di propaganda e disinformazione condotte sulla piattaforma *TikTok* grazie al sostegno di soggetti terzi riconducibili alla Russia avessero avuto l'effetto di manipolare la volontà degli elettori¹⁰.

La consapevolezza della facilità con cui le tecnologie digitali possano essere utilizzate per destabilizzare le istituzioni democratiche non implica che l'intelligenza artificiale debba essere demonizzata, o vietata *tout court*. Se è vero, come si è visto, che i sistemi di IA possano favorire la diffusione e persino la creazione di notizie false, è altrettanto vero che tali strumenti possano contribuire anche ad arginare il fenomeno della disinformazione, o perlomeno a ridurne le conseguenze dannose¹¹. Sotto questo profilo, ad esempio, i meccanismi di IA, attraverso sofisticati modelli di analisi del dato testuale, possono aiutare a rintracciare i contenuti potenzialmente falsi e ingannevoli pubblicati nella rete. Tale circostanza impone, a livello regolatorio, un'attenta ponderazione degli interessi in gioco, affinché, da un lato, non siano ostacolati lo sviluppo e la progettazione dei sistemi di IA, dall'altro, ne siano impediti eventuali utilizzi distorsivi per finalità discriminatorie, se non addirittura sovversive.

Si tratta, come appare evidente, di una sfida piuttosto complessa da affrontare, non solo perché presuppone, da parte del legislatore, un delicato bilanciamento tra esigenze contrapposte, ma soprattutto perché sconta una debolezza strutturale nel sistema dell'Unione europea, ossia l'assenza, in capo a quest'ultima, di una competenza esplicita in materia di disinformazione: una carenza che, vale la pena ricordarlo, non può essere superata per il tramite dell'art. 2 TUE, giacché tale disposizione non costituisce, come già evidenziato in

⁸ L. CALIFANO, F. FABBRIZZI, G. SARTOR, *Information disorder e sistema democratico*, in *federalismi.it*, n. 15, 2025, p. iv ss. In generale, sul tema, cfr. A. NICITA', *Il mercato delle verità: come la disinformazione minaccia la democrazia*, Bologna, 2024.

⁹ Comunicazione della Commissione, *Orientamenti della Commissione per i fornitori di piattaforme online di dimensioni molto grandi e di motori di ricerca online di dimensioni molto grandi sull'attenuazione dei rischi sistemici per i processi elettorali a norma dell'articolo 35, paragrafo 3, del regolamento (UE) 2022/2065*, 26 aprile 2024, C/2024/3014, par 36.

¹⁰ La vicenda ha suscitato ampio interesse sia a livello mediatico che accademico, poiché chiama in causa due aspetti essenziali per la tenuta degli ordinamenti democratici: da un lato, l'esigenza di salvaguardare il legittimo interesse di uno Stato sovrano di impedire ingerenze straniere nei propri affari interni senza, al contempo, svuotare di significato la sovranità popolare; dall'altro, la consapevolezza della fragilità dei sistemi democratici di fronte alle minacce poste dalle tecnologie digitali. Per un'analisi del caso rumeno, cfr., *ex multis*, D. A. CĂRĂMIDARIU, A. VERTEȘ-OLTEANU, *Safeguarding democracy: constitutional insights from Romania's election annulment*, in *Diritto pubblico comparato ed europeo*, 2025, p. 139 ss.; F. ROSA, *L'annullamento delle elezioni presidenziali in Romania e la difficile difesa della democrazia*, in *federalismi.it*, n. 15, 2025, p. 62 ss.; A. IANNOTTI DELLA VALLE, *Libertà di espressione e valori democratici alla prova dei social media: il DSA e un nuovo caso TikTok europeo*, in *federalismi.it*, n. 13, 2025, p. 84 ss.; B. SELEJAN-GUTAN, *The Second Round that Wasn't: Why The Romanian Constitutional Court Annulled the Presidential Elections*, in *Verfassungsblog*, 7 December 2024.

¹¹ Al riguardo, CONTISSA e GALLI hanno parlato di «disinformazione aumentata», intendendo, con tale espressione, «l'insieme delle trasformazioni che il fenomeno della disinformazione subisce in conseguenza dell'integrazione dell'IA nei processi di produzione, diffusione e gestione dei contenuti informativi. L'IA agisce come un potente moltiplicatore, capace sia di rendere la disinformazione più veloce, personalizzata e convincente, sia di offrire strumenti efficaci per contrastarla» (G. CONTISSA, F. GALLI, *La governance della "disinformazione aumentata" tra Digital Services Act e AI Act*, in *federalismi.it*, n. 15, 2025, p. 131).

dottrina, una base giuridica attributiva di competenza¹². Questi limiti, tuttavia, non hanno impedito all'Unione di intervenire nel settore *de quo* nel solco della c.d. via europea all'evoluzione tecnologica, ossia un modello di sviluppo che, pur incoraggiando la transizione digitale, non intende abdicare al substrato di valori condiviso a livello sovranazionale¹³.

È in questo scenario che si collocano le iniziative normative intraprese dall'Unione europea in materia di lotta alla disinformazione¹⁴, la quale, pur senza essere oggetto di una regolamentazione specifica, è venuta in rilievo nell'ambito di taluni strumenti legislativi adottati nel contesto del “decennio digitale europeo”¹⁵. Ci si riferisce, in particolar modo, al *Digital Services Act* (DSA)¹⁶ e all'*AI Act*¹⁷, i quali, sebbene caratterizzati da una matrice prevalentemente economica¹⁸, mirano a garantire che lo sviluppo e la gestione di specifici servizi e prodotti digitali nel mercato interno – sia pur nella sua dimensione di “mercato unico digitale” – non comporti un'erosione dei valori fondamentali dell'Unione¹⁹. Dal momento che, come si è già avuto modo di notare, la disinformazione può rappresentare un fattore di rischio per la solidità dello Stato di diritto e dei processi democratici, pilastri del sistema valoriale alla base dell'ordinamento dell'Unione, non sorprende che tale aspetto abbia trovato spazio negli atti normativi menzionati.

¹² A. ADINOLFI, *L'intelligenza artificiale tra rischi di violazione dei diritti fondamentali e sostegno alla loro promozione: considerazioni sulla (difficile) costruzione di un quadro normativo dell'Unione*, in A. PAJNO, F. DONATI, A. PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione? Diritti fondamentali, dati personali e regolazione*, vol. I, Bologna, 2022, p. 127 ss.

¹³ A. ADINOLFI, *Evoluzione tecnologica e tutela dei diritti fondamentali: qualche considerazione sulle attuali strategie normative dell'Unione*, in *I Post di AISDUE*, 14 marzo 2023, p. 321 ss.; F. FERRI, *Transizione digitale e valori fondanti dell'Unione: riflessioni sulla costituzionalizzazione dello spazio digitale europeo*, in *Il Diritto dell'Unione Europea*, 2022, p. 277 ss.

¹⁴ In generale sul contrasto alla disinformazione online nell'Unione europea, cfr. A. FESTA, *La disinformazione online come “minaccia ibrida” alla democrazia nell'Unione europea: meccanismi di tutela e strumenti a contrasto per uno spazio di libertà, sicurezza e giustizia*, in *Freedom, Security & Justice: European Legal Studies*, 2025, p. 272 ss.; J. BAYER, *The EU policy on disinformation: aims and legal basis*, in *Journal of Media Law*, n. 1, 2024, p. 18 ss.; S. SASSI, *L'Unione Europea e la lotta alla disinformazione online*, in *federalismi.it*, n. 3, 2023, p. 183 ss.

¹⁵ Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, *Bussola per il digitale 2030: il modello europeo per il decennio digitale*, 9 marzo 2021, COM(2021) 118 fin.

¹⁶ Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio del 19 ottobre 2022 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE, in *GUUE* L 277 del 27 ottobre 2022, p. 1 ss.

¹⁷ Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828, in *GUUE* L 1689 del 12 luglio 2024, p. 1 ss.

¹⁸ Ne è una riprova il fatto che entrambi i regolamenti – al pari, peraltro, di tutti gli altri adottati nell'ambito del mercato unico digitale, con la sola eccezione della direttiva 2024/2831 sul miglioramento delle condizioni di lavoro mediante piattaforme digitali, fondata sull'art. 153, par. 2, lett. b), TFUE – si basino sull'art. 114 TFUE relativo al ravvicinamento delle normative nazionali aventi ad oggetto l'instaurazione ed il funzionamento del mercato interno, cui si aggiunge in qualche caso, *AI Act* compreso, l'art. 16 TFUE per i profili riguardanti il trattamento dei dati personali. Sull'utilizzo di tali disposizioni per l'adozione di atti normativi nel settore digitale, cfr. S. POLI, *Il rafforzamento della sovranità tecnologica europea e il problema delle basi giuridiche*, in *I Post di AISDUE*, 20 dicembre 2021, p. 69 ss. In generale sul ricorso all'art. 114 TFUE come base giuridica, si rinvia a T. M. MOSCHETTA, *Il ravvicinamento delle normative nazionali per il mercato interno. Riflessioni sul sistema delle fonti alla luce dell'art. 114 TFUE*, Bari, 2018.

¹⁹ In tal senso, con riferimento all'*AI Act*, cfr. R. PALLADINO, *L'“approccio europeo” al contrasto alla disinformazione digitale e alla protezione dei valori democratici: quale contributo dell'AI Act?*, in *Freedom, Security & Justice: European Legal Studies*, n. 2, 2025, p. 296 ss.; F. M. MANCIOPPI, *La regolamentazione dell'intelligenza artificiale come opzione per la salvaguardia dei valori fondamentali dell'UE*, in *federalismi.it*, n. 7, 2024, p. 112 ss.

Il contrasto alla disinformazione costituisce, non a caso, un elemento centrale degli obblighi di *compliance* incombenti alle piattaforme digitali e ai motori di ricerca di dimensioni molto grandi ai sensi del DSA²⁰, nella convinzione che i *social network*, consentendo la rapida trasmissione dei contenuti a un numero potenzialmente enorme di utenti, possano essere utilizzati per «diffondere o amplificare contenuti fuorvianti o ingannevoli», suscettibili di produrre effetti negativi su questioni di interesse generale per la collettività, quali la salute pubblica, il pluralismo dei media e l'integrità delle consultazioni elettorali²¹. In ragione di ciò, secondo quanto sancito dall'art. 34, siffatte piattaforme sono tenute ad individuare, analizzare e valutare con diligenza i *rischi sistemici* derivanti dai loro servizi, compresi gli «eventuali effetti negativi, attuali o prevedibili, sul dibattito civico e sui processi elettorali, nonché sulla sicurezza pubblica», apportando, laddove necessario, adeguate misure correttive, le quali possono interessare, *inter alia*, il funzionamento dei relativi sistemi algoritmici e di raccomandazione²². Il ruolo del DSA nel contrastare la disinformazione nella sfera digitale è già stato ampiamente discusso in dottrina²³. In questa sede, pertanto, si intende porre l'attenzione sul solo *AI Act*²⁴, al fine di verificare se, e in che misura, tale regolamento contribuisca ad affrontare i rischi correlati alla disinformazione *online*²⁵.

2. Brevi cenni sulla struttura e sulle finalità dell'AI Act

Si è già osservato come l'approccio dell'Unione europea al progresso tecnologico intenda collocare gli sviluppi inerenti al completamento del mercato unico digitale entro un solido orizzonte valoriale, al fine di evitare che la progettazione, lo scambio e l'utilizzo (in

²⁰ Ai sensi dell'art. 33 del DSA, per piattaforme digitali e motori di ricerca di dimensioni molto grandi si intendono gli operatori con un bacino medio mensile di destinatari del servizio nell'Unione almeno pari a 45 milioni.

²¹ Regolamento (UE) 2022/2065, cit., considerando n. 84.

²² Per una panoramica di tali profili del DSA, cfr. L. D'AGOSTINO, *Disinformazione e obblighi di compliance degli operatori del mercato digitale alla luce del nuovo Digital Services Act*, in *MediaLaws*, n. 2, 2023, p. 33 ss.

²³ Cfr. E. LONGO, *Libertà di informazione e lotta alla disinformazione nel Digital Services Act*, in *Giornale di diritto amministrativo*, 2023, p. 737 ss.; A. STROWEL, J. DE MEYERE, *The Digital Services Act: transparency as an efficient tool to curb the spread of disinformation on online platforms?*, in *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, n. 1, 2023, p. 66 ss.; G. CAGGIANO, *Il contrasto alla disinformazione tra nuovi obblighi delle piattaforme online e tutela dei diritti fondamentali nel quadro del Digital Service Act e della co-regolamentazione*, in *Papers di diritto europeo*, 2021, p. 45 ss.

²⁴ L'*AI Act* è stato oggetto di numerosi studi, che ne hanno indagato i vari profili d'interesse. Per un inquadramento generale del regolamento e delle sue implicazioni più rilevanti sotto il profilo giuridico, cfr., *ex multis*, M. ALMADA, N. PETIT, *The EU AI Act: Between the Rock of Product Safety and the Hard Place of Fundamental Rights*, in *Common Market Law Review*, 2025, p. 850 ss.; F. DONATI, G. FINOCCHIARO, F. PAOLUCCI, O. POLLICINO, *La disciplina dell'intelligenza artificiale: questioni e prospettive*, in *Rivista Quaderni AISDUE*, fascicolo speciale n. 2, 2024; R. PETRUSO, G. SMORTO, *Il Regolamento europeo sull'intelligenza artificiale: una prima lettura*, in *La Nuova Giurisprudenza civile commentata*, 2024, p. 989 ss.; C. GRIECO, *Intelligenza Artificiale e tutela degli utenti nel diritto dell'Unione europea*, Napoli, 2023, spec. p. 59 ss.; N. T. NIKOLINAKOS, *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies – The AI Act*, Cham, 2023; G. CONTALDI, *La proposta di regolamento sull'intelligenza artificiale e la protezione di dati personali*, in G. CAGGIANO, G. CONTALDI, P. MANZINI (a cura di), *Verso una legislazione europea su mercati e servizi digitali*, Bari, 2022, p. 205 ss.

²⁵ Per una panoramica sul contributo dell'*AI Act* in materia di contrasto alla disinformazione, cfr. M. MESARČÍK, N. SLOSIAROVÁ, *Regulating AI for a truthful tomorrow: addressing disinformation in the EU Artificial Intelligence Act*, in *International Journal of Law and Information Technology*, vol. 33, 2025, p. 1 ss.

questo caso) dei sistemi di intelligenza artificiale possa seguire traiettorie eterogenee all'interno dei vari Paesi membri. Tale dialettica costituisce, invero, una prassi consolidata nell'azione legislativa dell'Unione, stante la consueta tendenza del legislatore europeo – riscontrabile soprattutto in materia di sicurezza dei prodotti – a far interagire le esigenze di armonizzazione del mercato interno con la protezione dei diritti fondamentali e di taluni interessi generali, quali l'ambiente e la salute pubblica, nell'ottica, da un lato, di realizzare quell'economia sociale di mercato invocata dall'art. 3, par. 3, TUE, dall'altro, di dare attuazione alle numerose clausole trasversali previste dai Trattati. Ciò detto, è chiaro che i pericoli posti dai sistemi di intelligenza artificiale, ancora parzialmente ignoti, impongano di rimodulare tale schema normativo, inquadrando la regolamentazione dell'IA secondo una prospettiva complessivamente orientata alla prevenzione e alla gestione del rischio²⁶.

Questa consapevolezza si è riflessa nella previsione di un articolato schema normativo, per cui le regole applicabili ai vari sistemi di intelligenza artificiale nell'ambito dell'*AI Act* non sono uniformi, bensì differiscono in ragione dei rischi che detti sistemi sono suscettibili di produrre sui diritti fondamentali e sui valori fondanti dell'Unione. Sebbene l'inclusione dei diritti fondamentali tra gli obiettivi di un atto normativo adottato dall'Unione non costituisca certo una novità, giacché il loro rispetto vincola le istituzioni europee nell'esercizio delle loro funzioni²⁷, l'*AI Act* inaugura una prospettiva inedita. In questo caso, infatti, lungi dal costituire – in continuità con la disciplina prevista abitualmente in materia di sicurezza dei prodotti – soltanto uno dei requisiti di cui tenere conto in fase di progettazione²⁸, i diritti fondamentali, nonché i valori di cui all'art. 2 TUE, configurano lo «scheletro» dell'intero *corpus* normativo²⁹, configurando il parametro cui associare la soglia di tollerabilità del rischio di un dato sistema di IA³⁰.

Abbracciando questa logica, il regolamento introduce, nei riguardi sia dei produttori che degli utilizzatori (*deployer*)³¹, una serie di divieti e obblighi procedurali di intensità variabile,

²⁶ Sotto questo profilo, con riferimento all'*AI Act*, cfr. M. INGLESE, *Il regolamento sull'intelligenza artificiale come atto per il completamento e il buon funzionamento del mercato interno*, in F. FERRI (a cura di), *op. cit.*, p. 71 ss.

²⁷ Come affermato dalla Corte nelle note sentenze *Ungheria* e *Polonia* sul meccanismo di condizionalità relativo allo Stato di diritto, l'art. 2 TUE, lungi dal costituire «una mera enunciazione di orientamenti o di intenti di natura politica [...] contiene valori che [...] fanno parte dell'identità stessa dell'Unione quale ordinamento giuridico comune, valori che sono concretizzati in principi che comportano obblighi giuridicamente vincolanti per gli Stati membri (sentenza della Corte (Seduta Plenaria) del 16 febbraio 2022, causa C-156/21, *Ungheria c. Parlamento e Consiglio*, ECLI:EU:C:2022:97, punto 232; sentenza della Corte (Seduta Plenaria) del 16 febbraio 2022, causa C-157/21, *Polonia c. Parlamento e Consiglio*, ECLI:EU:C:2022:98, punto 264).

²⁸ In tal senso, v. ad esempio il 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, in *GUUE* L 117 del 5 maggio 2017, p. 1 ss., il cui considerando n. 89 recita: «Il presente regolamento rispetta i diritti fondamentali e i principi riconosciuti, in particolare, dalla Carta e, segnatamente, la dignità umana, l'integrità della persona, la protezione dei dati di carattere personale, la libertà delle arti e delle scienze, la libertà d'impresa e il diritto di proprietà».

²⁹ E. CIRONE, *L'AI Act e l'obiettivo (mancato?) di promuovere uno standard globale per la tutela dei diritti fondamentali*, in F. FERRI (a cura di), *op. cit.*, p. 61.

³⁰ F. FERRI, *Transizione digitale e valori fondanti dell'Unione*, cit., p. 319.

³¹ Si tratta dei principali destinatari degli obblighi stabiliti dall'*AI Act*. In particolare, per fornitore si intende «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»; mentre il *deployer* è «una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che utilizza un sistema di IA sotto la propria autorità, tranne nel caso

a seconda della pericolosità attesa di ciascun sistema di IA. In sostanza, i sistemi di IA sono classificati in categorie distinte sulla base del livello di rischio che li caratterizza. Tale operazione presuppone un'analisi *ex ante* degli interessi in rilievo, allo scopo di valutare se le possibili conseguenze negative di un'applicazione di intelligenza artificiale siano tali da giustificare restrizioni più o meno rigorose alla loro immissione sul mercato o al loro utilizzo³².

Com'è facile immaginare, un impianto normativo così concepito non può tollerare lo sviluppo di sistemi di IA palesemente contrari ai diritti previsti dalla Carta e ai valori di cui all'art. 2 TUE, giacché nessuna misura correttiva adottata *ex post* sarebbe idonea a compensarne gli effetti dannosi. Nei confronti di tali sistemi, i cui rischi per la salute, la sicurezza e i diritti fondamentali sono ritenuti "inaccettabili", vige, pertanto, un divieto assoluto, sancito dall'art. 5 del regolamento 2024/1689. A un livello inferiore si situano i sistemi ad alto rischio, la cui regolazione costituisce la parte più significativa della disciplina introdotta dall'*AI Act*. Tale categoria comprende quei sistemi il cui livello di pericolosità, pur non essendo sufficiente a determinarne la proibizione, è tale da far sorgere, in capo ai fornitori, ai distributori e agli utilizzatori, specifici obblighi di *compliance*, al fine di minimizzarne i rischi attesi. Le disposizioni meno stringenti, di fatto limitate a meri obblighi di trasparenza e informazione, cui si aggiunge l'adesione volontaria a codici di condotta, si applicano, invece, ai sistemi a rischio minimo o limitato. Una disciplina a sé stante è riservata, infine, ai sistemi di IA "per scopi generali", i quali, potendo trovare applicazione per una molteplicità di compiti e funzioni, potrebbero comportare rischi sistemici, compresi, *inter alia*, «eventuali effetti negativi, effettivi o ragionevolmente prevedibili, sui processi democratici e sulla sicurezza pubblica ed economica», nonché «la diffusione di contenuti illegali, mendaci o discriminatori»³³.

Viste le finalità del presente scritto, occorre verificare in che modo tale approccio basato sul rischio possa venire in rilievo in materia di contrasto alla disinformazione posto che, come recita il considerando n. 133, «i nuovi rischi di cattiva informazione e manipolazione su vasta scala» costituiscono una delle minacce cui l'*AI Act* intende far fronte.

3. Le pratiche vietate ai sensi dell'art. 5: quale rilievo ai fini del contrasto alla disinformazione online?

Seguendo la struttura del regolamento in esame, il primo livello su cui occorre soffermarsi riguarda le pratiche vietate ai sensi dell'art. 5. Per i nostri fini rileva, nello specifico, l'ipotesi prevista dal par. 1, lett. a), la quale vieta «l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che utilizza tecniche subliminali che agiscono senza che una persona ne sia consapevole o tecniche volutamente manipolative o ingannevoli aventi lo scopo o l'effetto di distorcere materialmente il comportamento di una persona o di un gruppo di persone, pregiudicando in modo considerevole la loro capacità di prendere una decisione informata, inducendole pertanto a prendere una decisione che non avrebbero altrimenti preso, in un modo che provochi o possa ragionevolmente provocare a tale persona,

in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale» (regolamento (UE) 2024/1689, cit., art. 3, par. 3 e 4).

³² Su tali aspetti, cfr. C. SCHEPISI, *Le "dimensioni" della regolazione dell'intelligenza artificiale nella proposta di regolamento della Commissione*, in *I Post di AISDUE*, 28 marzo 2022, p. 330 ss., spec. p. 339 ss.

³³ Regolamento (UE) 2024/1689, cit., considerando n. 110.

a un'altra persona o a un gruppo di persone un danno significativo». La fattispecie sembrerebbe includere *prima facie* i sistemi di raccomandazione utilizzati dalle piattaforme digitali, nel caso in cui tali sistemi, favorendo la circolazione di contenuti disinformativi, possano indurre gli utenti in errore. Viene da chiedersi, pertanto, se siffatti sistemi rientrino nell'ambito di applicazione della disposizione in parola, dal momento che la norma non proibisce le tecniche subliminali *tout court*, ma soltanto quelle che rispondano alle caratteristiche espressamente previste.

Innanzitutto, occorre evidenziare che la disposizione distingue tra “tecniche subliminali” e “tecniche volutamente manipolative o ingannevoli”, senza fornire, tuttavia, indicazioni utili a comprendere in cosa consistano tali applicazioni di IA. Per quanto riguarda le tecniche subliminali, un ausilio a fini interpretativi perviene dal considerando n. 29 del regolamento 2024/1689, il quale menziona le tecniche di manipolazione basate sull'IA che impiegano «stimoli audio, grafici e video che le persone non sono in grado di percepire poiché tali stimoli vanno al di là della percezione umana o altre tecniche manipolative o ingannevoli che sovvertono o pregiudicano l'autonomia, il processo decisionale o la libera scelta di una persona senza che sia consapevole di tali tecniche o, se ne è consapevole, senza che sia in grado di controllarle o resistervi o possa evitare l'inganno». La definizione sembrerebbe riferirsi, in sostanza, a tutti i sistemi di IA in grado di influenzare le decisioni e i comportamenti dei soggetti in modo subdolo, ossia senza che questi ultimi ne abbiano coscienza, compromettendone la capacità di agire autonomamente e consapevolmente.

Appare invece meno agevole capire cosa contraddistingua le “tecniche volutamente manipolative o ingannevoli”. L'espressione sembra porre l'enfasi sull'intenzione del produttore (o utilizzatore) di un sistema di IA di ingannare deliberatamente gli utenti, sfruttandone i possibili fattori di vulnerabilità con l'obiettivo di influenzarne o controllarne il comportamento. Ciò posto, la distinzione rispetto alle semplici tecniche subliminali rimane ambigua, dal momento che, alla luce della formulazione del già citato art. 5, la volontà del fornitore o del *deployer* di distorcere il comportamento delle persone non rileva al fine di accertare l'illiceità di una determinata pratica, essendo sufficiente provare il compimento dell'effetto manipolatorio. In effetti, come precisato nelle linee guida pubblicate dalla Commissione sulle pratiche di intelligenza artificiale vietate ai sensi dell'*AI Act*³⁴, ciò che accomuna le varie tecniche proibite risiede nel risultato che esse intendono perseguire, ossia indurre gli utenti a tenere un comportamento o a prendere una decisione che altrimenti non avrebbero assunto.

Quest'ultimo rappresenta l'aspetto più controverso della disposizione *de qua*, poiché presuppone che il comportamento umano in questione costituisca una diretta conseguenza della manipolazione effettuata tramite l'utilizzo dell'intelligenza artificiale. Pertanto, nel caso in cui l'IA si limitasse a condizionare in modo generico l'operato dei soggetti, il divieto *ex art.* 5 non troverebbe applicazione, dovendosi riscontrare, viceversa, «un impatto sostanziale sul comportamento» di una persona tale da comprometterne l'autonomia e la libertà di scelta³⁵. Detto altrimenti, soltanto l'esistenza di un nesso causale tra la tecnica subliminale utilizzata e l'alterazione del comportamento umano renderebbe illegittima siffatta applicazione di IA. Si tratta, come risulta evidente, di un punto piuttosto problematico della disposizione in parola, suscettibile di restringerne sensibilmente la sfera d'efficacia.

³⁴ Comunicazione della Commissione, *Orientamenti della Commissione relativi alle pratiche di intelligenza artificiale vietate ai sensi del regolamento (UE) 2024/1689 (regolamento sull'IA)*, 29 luglio 2025, C(2025) 5052 fin.

³⁵ *Ibidem*, par. 76.

Con riferimento ai profili d'interesse in questa sede, il divieto si estenderebbe ai sistemi di IA in grado di incidere in modo significativo sulle preferenze di voto degli elettori. Al riguardo, vale la pena soffermarsi sulla vicenda rumena evocata in precedenza. In quel caso, infatti, la Corte costituzionale ha sostenuto che la diffusione *online* di contenuti disinformativi su tematiche di rilievo politico (di cui era peraltro taciuta la matrice elettorale), agevolata dai sistemi di raccomandazione di TikTok, avesse impedito agli elettori di ricevere informazioni puntuali sui candidati, così da formarsi liberamente un'opinione senza essere influenzati in modo illecito e sproporzionato. Tale ragionamento sembrava sottintendere una relazione diretta tra la campagna disinformativa condotta sfruttando gli algoritmi della piattaforma e il risultato elettorale.

Ciò detto, non si può non considerare che la decisione della Corte costituzionale sia maturata in uno scenario di grande incertezza politica, giacché, appena due giorni prima della pronuncia, un *dossier* dei servizi di *intelligence* aveva rilevato interferenze russe durante la campagna elettorale. Tale circostanza ha indotto il giudice costituzionale ad intervenire *ex officio* prima che si svolgesse il secondo turno delle elezioni – in programma quattro giorni dopo la pubblicazione del *dossier* – penalizzando la solidità dell'argomentazione giuridica³⁶. Pur costituendo un precedente di rilievo, la vicenda non consente, proprio per le circostanze specifiche che l'hanno accompagnata, di pervenire a facili conclusioni. A livello generale, infatti, ci sembra improbabile che la divulgazione di informazioni false possa configurare un fattore di distorsione del comportamento nell'accezione accolta dall'art. 5, par. 1, lett. a), dell'*AI Act*. Del resto, pure di fronte a un'intensa strategia disinformativa, sarebbe possibile sostenere che, in assenza di tali azioni intrusive, il risultato elettorale sarebbe stato diverso? Per quanto verosimile in linea teorica, rimane un'ipotesi difficile da dimostrare con certezza, o, quantomeno, con un certo grado di plausibilità.

È pur vero, come suggeriscono le linee guida della Commissione³⁷, che la nozione di «distorsione materiale del comportamento» possa essere interpretata in parallelo agli effetti distorsivi causati dalle pratiche commerciali sleali, ingannevoli e aggressive di cui agli artt. 5-9 della direttiva 2005/29/CE³⁸ nei confronti dei consumatori. Orbene, come precisato dalla Corte di giustizia, per essere ritenuta ingannevole è sufficiente che una pratica *sia idonea* a indurre il consumatore ad assumere una decisione commerciale che altrimenti non avrebbe preso, anche qualora il comportamento del soggetto non sia stato effettivamente falsato nel caso specifico³⁹. A tal proposito, peraltro, la Corte ha avuto modo di evidenziare le criticità derivanti dalle pratiche che mirano a «sfruttare l'effetto psicologico creato nella mente del consumatore [...]»⁴⁰. Tale aspetto appare particolarmente significativo nel caso dei sistemi di

³⁶ Oltre a ciò, la sentenza ha ricevuto critiche, sotto il profilo procedurale, per l'intervento *ex officio* anziché su istanza di parte, per le ristrette tempistiche con cui la procedura si è svolta e per il mancato coinvolgimento delle parti interessate. Su questo dibattito, si rinvia a F. ROSA, *L'annullamento delle elezioni presidenziali in Romania e la difficile difesa della democrazia*, cit.

³⁷ *Orientamenti della Commissione relativi alle pratiche di intelligenza artificiale vietate*, cit., par. 80 ss.

³⁸ Direttiva 2005/29/CE del Parlamento europeo e del Consiglio, dell'11 maggio 2005, relativa alle pratiche commerciali sleali tra imprese e consumatori nel mercato interno e che modifica la direttiva 84/450/CEE del Consiglio e le direttive 97/7/CE, 98/27/CE e 2002/65/CE del Parlamento europeo e del Consiglio e il regolamento (CE) n. 2006/2004 del Parlamento europeo e del Consiglio, in *GUCE* L 149 dell'11 giugno 2005, p. 22 ss.

³⁹ Sentenza della Corte di giustizia (Quinta Sezione) del 26 ottobre 2016, *Canal Digital Danmark A/S*, causa C-611/14, ECLI:EU:C:2016:800, punto 37.

⁴⁰ Sentenza della Corte (Sesta Sezione) del 18 ottobre 2012, *Purely Creative e a.*, causa C-428/11, ECLI:EU:C:2012:651, punto 49.

raccomandazione utilizzati dalle piattaforme, i quali, come ricordato in precedenza, fanno leva sulla partecipazione emotiva degli utenti, valorizzando i contenuti più coinvolgenti, suscettibili di plasmare le percezioni degli individui senza che ne abbiano consapevolezza. Questi sviluppi giurisprudenziali in materia di protezione dei consumatori offrono un riferimento cui la Corte potrebbe attingere qualora fosse chiamata, ad esempio nell'ambito di un rinvio pregiudiziale *ex art. 267 TFUE*, a chiarire la nozione di «distorsione materiale del comportamento» ai sensi dell'art. 5, par. 1, lett. *a*). Tuttavia, dal momento che le preferenze di voto sono spesso volubili, rimarrebbe comunque complesso dimostrare l'esistenza di un nesso causale tra la pratica di IA impiegata e l'alterazione del comportamento elettorale.

In ogni caso, anche qualora un simile nesso causale fosse verificabile, i sistemi di raccomandazione delle piattaforme rientrerebbero difficilmente nel perimetro applicativo dell'art. 5 del regolamento 2024/1689. Ciò perché, secondo il dettato di detta disposizione, la causazione di un «danno significativo», o, quantomeno, la circostanza per cui siffatto scenario risulti «ragionevolmente» prevedibile, rappresenta una *conditio sine qua non* affinché una specifica tecnica manipolatoria basata sull'IA debba essere vietata. Il già citato considerando n. 29 precisa che per «danni significativi» debbano intendersi gli «effetti negativi sufficientemente importanti sulla salute fisica, psicologica o sugli interessi finanziari». Per i nostri fini, l'unica tipologia potenzialmente rilevante è costituita dai danni di tipo psicologico, in quanto correlati ai sistemi di intelligenza artificiale «che sfruttano le vulnerabilità cognitive ed emotive e influenzano i comportamenti di una persona»⁴¹. Come precisano le linee guida pubblicate dalla Commissione, tali danni comprendono gli «effetti negativi sulla salute mentale e sul benessere psicologico ed emotivo di una persona»⁴²: sebbene la circolazione di contenuti disinformativi possa inficiare la capacità delle persone di distinguere il vero dal falso, ci sembra difficile che l'applicazione dell'IA per scopi disinformativi possa comportare un danno nel senso inteso dalla disposizione in parola⁴³.

Ricapitolando, vista la necessità di provare la contestuale presenza dei requisiti analizzati, appare improbabile che l'utilizzo dell'intelligenza artificiale per finalità di disinformazione – specialmente tramite i sistemi di raccomandazione – possa configurare una pratica vietata ai sensi dell'art. 5 dell'*AI Act*⁴⁴.

4. Il contrasto alla disinformazione nell'ambito dei sistemi di intelligenza artificiale ad alto rischio

Ai fini del contrasto alla disinformazione *online*, vale la pena soffermarsi anche sulla regolamentazione riservata ai sistemi di intelligenza artificiale ad alto rischio di cui all'art. 6 dell'*AI Act*. Come ricordato in precedenza, la disciplina di tali sistemi costituisce la parte più significativa dell'approccio basato sul rischio introdotto dal regolamento 2024/1689. Le

⁴¹ *Orientamenti della Commissione relativi alle pratiche di intelligenza artificiale vietate*, cit., par. 88.

⁴² *Ibidem*.

⁴³ Si noti che, durante la fase di discussione della proposta di regolamento sull'IA presentata dalla Commissione, il Comitato economico e sociale europeo ha suggerito di includere, tra le pratiche vietate *ex art. 5*, par. 1, lett. *a*), quelle che causano «pregiudizio ai diritti fondamentali, compresi quelli alla salute e alla sicurezza fisiche o psicologiche, di un'altra persona o di un gruppo di persone, o alla democrazia e allo Stato di diritto» (parere del Comitato economico e sociale europeo sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale e modifica alcuni atti legislativi dell'Unione, COM(2021) 206 fin., 2021/0106 (COD), par. 4.3). Tale modifica, tuttavia, non è stata accolta dal legislatore.

⁴⁴ In tal senso, M. MESARČIK, N. SLOSIAROVÁ, *Regulating AI for a truthful tomorrow*, cit., p. 8.

applicazioni di IA ricomprese in questa categoria presentano rischi elevati per la salute, la sicurezza e i diritti fondamentali: ragion per cui l'immissione sul mercato unico, la messa in servizio e l'utilizzo di tali applicazioni sono subordinati al rispetto di determinati requisiti tecnici, cui si accompagnano specifici obblighi di *compliance* posti in capo ai fornitori e agli utilizzatori. Il regolamento non intende bandire siffatti sistemi dal mercato, bensì ridurne i rischi potenziali, cosicché la loro progettazione e distribuzione non comporti pericoli eccessivi per gli individui e la collettività.

Per quanto d'interesse in questa sede, rileva il contenuto dell'allegato III, il quale, dando seguito all'art. 6, par. 2, del regolamento *de quo*, elenca, distinguendoli per settore di applicazione, una serie di sistemi di IA ritenuti ad alto rischio⁴⁵. La sezione dedicata all'amministrazione della giustizia e ai processi democratici menziona, *inter alia*, i sistemi «destinati a essere utilizzati per influenzare l'esito di un'elezione o di un referendum o il comportamento di voto delle persone fisiche nell'esercizio del loro voto alle elezioni o ai referendum»⁴⁶. La qualificazione di tali sistemi come ad alto rischio deriva dalla necessità di prevenire eventuali effetti negativi sulla democrazia e sullo Stato di diritto, nonché «i rischi di indebite interferenze esterne sul diritto di voto» sancito dall'art. 39 della Carta⁴⁷.

Si noti, tuttavia, che la riconducibilità di uno strumento di IA a una delle ipotesi descritte dall'allegato III non ne determina *ipso facto* la classificazione come sistema ad alto rischio, giacché, per effetto dell'art. 6, par. 3, tale dicitura è limitata ai sistemi che presentino un «rischio significativo di danno» per la salute, la sicurezza o i diritti fondamentali delle persone fisiche⁴⁸, valutato alla luce della gravità del possibile danno, nonché della probabilità che esso si verifichi⁴⁹. Si tratta, come appare evidente, di un criterio piuttosto sfuggente, giacché la valutazione di «significatività» non si basa sulla gravità di un danno effettivo, bensì sulla mera prevedibilità di quest'ultimo. Ne consegue che i fornitori dispongano di un apprezzabile margine di discrezionalità nel valutare se il loro sistema rientri nell'ambito di applicazione della deroga *ex art.* 6, par. 3, dal momento che la decisione di un fornitore di invocare tale clausola, pur dovendo essere documentata, non è soggetta ad alcun controllo esterno da parte di un'autorità di vigilanza⁵⁰.

Venendo al regime normativo applicabile, le tecniche di IA suscettibili di interferire con i processi elettorali sono soggette, al pari degli altri sistemi inclusi nella fascia di rischio

⁴⁵ Ai sensi dell'art. 6, par. 1, sono inoltre considerati ad alto rischio i sistemi di IA destinati a essere utilizzati come componente di sicurezza di un prodotto, o che costituiscano essi stessi un prodotto, già disciplinati dalla normativa di armonizzazione dell'Unione elencata nell'allegato I.

⁴⁶ Regolamento (UE) 2024/1689, cit., allegato III, par. 8, lett. b). Rimangono esclusi dalla qualifica di sistemi ad alto rischio i «sistemi di IA ai cui output le persone fisiche non sono direttamente esposte, come gli strumenti utilizzati per organizzare, ottimizzare e strutturare le campagne politiche da un punto di vista amministrativo e logistico».

⁴⁷ *Ibidem*, considerando n. 62.

⁴⁸ La disposizione precisa che un sistema non possa essere ritenuto ad alto rischio se *non* influenza «materialmente il risultato del processo decisionale»; viceversa, «è sempre considerato ad alto rischio qualora esso effettui profilazione di persone fisiche». Per una lettura critica del concetto di «danno» come soglia per la protezione dei diritti fondamentali, cfr. L. GROZDANOVSKI, J. DE COOMAN, *Forget the Facts, Aim for the Rights! On the Obsolescence of Empirical Knowledge in Defining the Risk/Rights-Based Approach to AI Regulation in the European Union*, in *Rutgers Computer and Technology Law Journal*, n. 2, 2023, p. 207 ss.

⁴⁹ Regolamento (UE) 2024/1689, cit., considerando n. 52.

⁵⁰ *Ibidem*, art. 6, par. 4. A tal proposito, si ricorda che, in linea generale, l'immissione sul mercato dei sistemi di IA ad alto rischio non è subordinata al rilascio di una licenza o a una forma di autorizzazione preventiva, bensì avviene sulla base di un'autovalutazione *ex ante* da parte del fornitore volta ad accertare la conformità del sistema da esso sviluppato ai requisiti tecnici stabiliti dal regolamento (art. 43).

elevato, al sistema di gestione dei rischi di cui all'art. 9 del regolamento 2024/1689. Tale quadro normativo, evidentemente mutuato dal meccanismo di gestione dei rischi sistemici istituito dagli artt. 34-35 del DSA, impone ai fornitori di identificare i potenziali rischi per la salute, la sicurezza e i diritti fondamentali derivanti dall'impiego del sistema di IA da essi progettato, nonché di apportare, se necessario, le opportune misure correttive⁵¹. Oltre a ciò, i fornitori dei sistemi di IA in questione sono tenuti, prima della loro immissione sul mercato, a valutarne la conformità rispetto a specifici *standard* di sicurezza, nonché a prevedere adeguati meccanismi di supervisione umana in grado di scongiurare eventuali usi impropri ed effetti indesiderati.

Anche i *deployer*, ossia gli utilizzatori di una tecnologia di IA, siano essi organismi di diritto pubblico o enti privati che forniscono servizi pubblici – nel nostro caso, ad esempio, i partiti politici o le agenzie di comunicazione che organizzano le campagne elettorali – sono destinatari di obblighi specifici. In particolare, ai sensi dell'art. 27 dell'*AI Act*, tali soggetti sono tenuti a valutare se il sistema di intelligenza artificiale da essi impiegato possa arrecare un pregiudizio ai diritti fondamentali⁵². Logicamente, i diritti fondamentali maggiormente attenzionati varieranno di volta in volta in ragione del settore in cui opera il sistema di IA implementato. Va da sé che, con riferimento ai sistemi volti ad influenzare l'andamento delle consultazioni elettorali, la valutazione d'impatto dovrà accertare se l'impiego di quel sistema comporti rischi significativi soprattutto in termini di pregiudizio alla libertà di espressione, alla tutela della privacy e al diritto a un'informazione libera e pluralistica. È in questa sede, pertanto, che i pericoli correlati all'utilizzo dell'intelligenza artificiale per scopi disinformativi dovrebbero essere mitigati.

Ci sembra ancora prematuro stabilire se uno schema normativo così delineato sia in grado di affrontare efficacemente i rischi per la democrazia derivanti dalle strategie disinformative messe in campo tramite l'IA. Tuttavia, non si può fare a meno di osservare che il regolamento sconti, sotto questo profilo, taluni vizi strutturali già segnalati in dottrina⁵³. In particolare, occorre domandarsi se la prospettiva che ispira l'*AI Act*, basata sull'accettabilità del rischio quale criterio di riferimento, sia funzionale ad arginare il fenomeno della disinformazione *online*. Come già accennato, il regolamento muove dalla premessa secondo cui il rischio altro non sia che la combinazione della probabilità ed entità dell'eventuale danno: di conseguenza, affinché il livello di rischio atteso sia tollerabile, è necessario che tali fattori rispettino uno *standard* minimo accettabile. Una simile impostazione – debitrice degli strumenti di valutazione del rischio elaborati nell'ambito della legislazione in materia di sicurezza dei prodotti – presuppone che il danno eventuale risulti misurabile e prevedibile attraverso modelli statistici, nonché univoco, nel senso di costituire l'unico possibile effetto indesiderato correlato a un determinato rischio.

Difficilmente siffatto schema potrebbe rivelarsi idoneo a contrastare efficacemente i pericoli per la democrazia connessi all'utilizzo dei sistemi di IA per ragioni disinformative. La capacità di un contenuto disinformativo di condizionare in via autonoma i processi

⁵¹ Per un'analisi di tali profili del DSA, cfr. E. BIRITTERI, *Contrasto alla disinformazione, Digital Services Act e attività di private enforcement: fondamento, contenuti e limiti degli obblighi di compliance e dei poteri di autonormazione degli operatori*, in *MediaLaws*, n. 2, 2023, p. 52 ss.

⁵² Per approfondire la valutazione d'impatto sui diritti fondamentali prevista dall'*AI Act*, cfr. A. MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template*, in *Computer Law & Security Review*, vol. 54, 2024, p. 1 ss.

⁵³ Cfr. M. ALMADA, N. PETIT, *The EU AI Act*, cit., p. 103 ss.; E. CIRONE, *L'AI Act e l'obiettivo (mancato?) di promuovere uno standard globale per la tutela dei diritti fondamentali*, cit., p. 61 ss.

elettorali, dunque senza l'interazione di altri eventi esterni, risulta, infatti, piuttosto difficile da prevedere *a priori*, vista l'incertezza che accompagna solitamente le intenzioni di voto degli individui. Inoltre, gli effetti negativi causati dalla disinformazione *online* potrebbero concretizzarsi solamente nel lungo periodo, peraltro attraverso manifestazioni non facilmente tangibili. Ad esempio, la circolazione di informazioni manipolate o fuorvianti su questioni di interesse politico-sociale potrebbe indebolire la fiducia delle persone nei confronti delle istituzioni, minando le fondamenta dei sistemi democratici⁵⁴. In queste circostanze, l'approccio adottato dall'*AI Act* non impedirebbe la realizzazione dell'evento dannoso, poiché quest'ultimo non sarebbe riconducibile a una specifica applicazione dell'intelligenza artificiale, bensì costituirebbe l'esito di un articolato processo di interazione uomo-macchina⁵⁵.

5. Gli obblighi di trasparenza per i sistemi di IA a rischio minimo e le disposizioni specifiche in materia di deep fake

Alla base della struttura piramidale predisposta dall'*AI Act* si situano i sistemi di intelligenza artificiale a rischio minimo, i quali, ai sensi dell'art. 50, sono sottoposti soltanto ad obblighi di trasparenza. Al di là delle varie ipotesi specifiche previste da detta disposizione, la normativa risponde, nel complesso, a una precisa finalità: rendere i contenuti e gli strumenti basati sull'intelligenza artificiale facilmente riconoscibili agli utenti che si trovassero ad interagirvi. A tal proposito, l'art. 50, par. 1, del regolamento 2024/1689 obbliga i fornitori dei sistemi destinati a interagire direttamente con le persone fisiche a progettare e sviluppare tali sistemi in modo tale che queste ultime «siano informate del fatto di stare interagendo con un sistema di IA, a meno che ciò non risulti evidente dal punto di vista di una persona fisica ragionevolmente informata, attenta e avveduta, tenendo conto delle circostanze e del contesto di utilizzo». Come osservato in dottrina, tale deroga introduce un elemento di flessibilità potenzialmente suscettibile di limitare la portata degli obblighi informativi incombenti ai fornitori⁵⁶.

Per i nostri fini, in ogni caso, più delle applicazioni di IA volte ad interfacciarsi direttamente con gli utenti (i *chatbot*, ad esempio), a suscitare interesse sono soprattutto i contenuti generati tramite IA. I sistemi di intelligenza artificiale, infatti, in quanto capaci di riprodurre fedelmente luoghi e personaggi, possono essere utilizzati per raffigurare in modo verosimile situazioni in realtà mai accadute, ingannando gli utenti sull'autenticità degli eventi rappresentati: logicamente, più un contenuto realizzato tramite l'ausilio dell'IA appare realistico, più sarà difficile per i soggetti riconoscerne l'origine artificiale. Se impiegati in contesti politico-elettorali, tali strumenti, impattando negativamente «sull'integrità e sulla fiducia [delle persone] nell'ecosistema dell'informazione»⁵⁷, possono trasformarsi in potenti mezzi di manipolazione dell'opinione pubblica. Proprio in virtù di ciò, è fondamentale che i soggetti abbiano consapevolezza della natura autentica o artefatta dei contenuti con cui si

⁵⁴ Al riguardo, per quanto concerne i *deep fake*, cfr. R. CHESNEY, D. K. CITRON, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, in *California Law Review*, 2019, p. 1753 ss.

⁵⁵ M. ALMADA, N. PETIT, *The EU AI Act*, cit., p. 108.

⁵⁶ R. PALLADINO, L'«*approccio europeo*» al contrasto alla disinformazione digitale e alla protezione dei valori democratici, cit., p. 313.

⁵⁷ Regolamento (UE) 2024/1689, cit., considerando n. 133.

interfacciano. A tal fine, i fornitori dei sistemi di IA che generano contenuti audio, immagini, video o testuali sintetici sono tenuti a marcare «in un formato leggibile meccanicamente» gli *output* prodotti dal proprio sistema, così da consentire di rilevare se questi ultimi siano stati «generati o manipolati artificialmente»⁵⁸.

Specificata attenzione è riservata, inoltre, ai *deep fake*⁵⁹. Secondo la nozione accolta dall'*AI Act*, l'espressione si riferisce a «un'immagine o un contenuto audio o video generato o manipolato dall'IA che assomiglia a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona»⁶⁰. Come spiegato dal Garante per la protezione dei dati personali, si tratta, in sostanza, di contenuti multimediali generati tramite sofisticati *software* di intelligenza artificiale capaci, attingendo a immagini e audio reali, di «modificare o ricreare, in modo estremamente realistico, le caratteristiche e i movimenti di un volto o di un corpo e a imitare fedelmente una determinata voce»⁶¹, cosicché una persona sembri pronunciare parole o compiere azioni che, in realtà, non ha mai detto o fatto. Ciò rende tali contenuti particolarmente efficaci in contesti elettorali, in quanto possono essere creati allo scopo di screditare o, al contrario, promuovere un determinato personaggio politico, proponendo narrazioni del tutto manipolate. Assicurare la riconoscibilità dei *deep fake* da parte degli utenti non risponde, pertanto, soltanto a un generico requisito di trasparenza, bensì anche alla necessità di proteggere il funzionamento dei sistemi democratici, evitando storture nel sistema dell'informazione in grado di alterare i meccanismi di consolidamento del consenso.

Al riguardo, l'art. 50, par. 4, primo comma, dell'*AI Act*, obbliga gli utilizzatori di un sistema di IA in grado di realizzare *deep fake* a rendere noto che «il contenuto è stato generato o manipolato artificialmente»⁶²: tuttavia, qualora il contenuto in questione «faccia parte di

⁵⁸ *Ibidem*, art. 50, par. 2.

⁵⁹ Per una prospettiva d'insieme sul fenomeno dei *deep fake* e sulla sua regolamentazione nell'ordinamento dell'Unione, cfr. M. J. BLOCK, *A Critical Evaluation of Deepfake Regulation through the AI Act in the European Union*, in *Journal of European Consumer and Market Law*, n. 4, 2024, p. 184 ss.; G. PROIETTI, *L'impianto regolatorio della società dell'informazione tra vecchi e nuovi equilibri. Il fenomeno del deep fake*, in *MediaLaws*, 2024, p. 328 ss.; A. RUFFO, *Il disordine informativo e l'Intelligenza Artificiale: tra insidie e possibili strumenti di contrasto*, in *MediaLaws*, 2024, p. 407 ss.; M. CAZZANIGA, *Una nuova tecnica (anche) per veicolare disinformazione: le risposte europee ai deepfakes*, in *MediaLaws*, 2023, p. 170 ss.

⁶⁰ Regolamento (UE) 2024/1689, cit., art. 3, par. 60. Secondo CONTISSA e GALLI, tale definizione solleva talune perplessità interpretative. In particolare, essa non chiarisce quale sia il livello di manipolazione richiesto affinché un contenuto possa essere classificato come *deep fake*: «ad esempio, nel caso delle immagini, posto che un'immagine scattata da una fotocamera digitale non mostra mai la realtà (vera), ma è il risultato della catena di elaborazione algoritmica dell'immagine e delle impostazioni scelte dal fotografo, la questione è se potrebbe basterebbe l'applicazione sull'immagine di un qualsiasi filtro digitale basato su AI (diffusissimi nei comuni smartphone) perché essa sia considerata manipolata, e quindi un deepfake, qualora ricorressero anche le altre condizioni» (G. CONTISSA, F. GALLI, *La governance della "disinformazione aumentata" tra Digital Services Act e AI Act*, cit., p. 150). Anche il concetto di somiglianza non appare sufficientemente nitido: ci si chiede, al riguardo, se occorra riscontrare la somiglianza di un contenuto con un elemento effettivamente esistente, oppure se basti rilevare il *realismo* di una figura, a prescindere che essa coincida o meno con un soggetto/oggetto realmente esistente. Per approfondire tali profili, si rinvia a K. MEDING, C. SORGE, *What Constitutes a Deep Fake? The Blurry Line Between Legitimate Processing and Manipulation Under the EU AI Act*, in *Proceedings of the 2025 Symposium on Computer Science and Law*, 2025, p. 152 ss.

⁶¹ Garante per la protezione dei dati personali, *Deepfake. Il falso che ti «rub» la faccia e la privacy*, 28 dicembre 2020.

⁶² Il medesimo obbligo si applica nei confronti del *deployer* di un sistema di IA «che genera o manipola testo pubblicato allo scopo di informare il pubblico su questioni di interesse pubblico», a meno che il contenuto generato dall'IA sia stato sottoposto «a un processo di revisione umana o di controllo editoriale e una persona fisica o giuridica detiene la responsabilità editoriale della pubblicazione del contenuto» (regolamento (UE) 2024/1689, cit., art. 50, par. 4, secondo comma).

un'analogia opera o di un programma manifestamente artistici, creativi, satirici o fittizi», si applica un obbligo di trasparenza meno rigoroso, poiché limitato a rivelare l'esistenza di tale contenuto generato artificialmente, «senza ostacolare l'esposizione o il godimento dell'opera». Viste le finalità del presente scritto, vale la pena soffermarsi su tale disposizione.

Innanzitutto, occorre evidenziare che il legislatore abbia preferito non disciplinare in modo eccessivamente stringente il fenomeno dei *deep fake*. I sistemi di IA in grado di realizzarli, infatti, non sono stati né proibiti, né inseriti tra i sistemi ad alto rischio, nonostante tali contenuti possano comportare evidenti pericoli in termini di mistificazione della realtà. A nostro avviso, un generale divieto di creazione e divulgazione dei *deep fake* sarebbe stato eccessivo, giacché, proibendo qualsiasi utilizzo di tali tecnologie, ne avrebbe impedito anche eventuali applicazioni positive⁶³. Del resto, ad apparire controversa non è la tecnologia in sé, di fatto neutrale, bensì gli scopi per la quale essa viene implementata⁶⁴. Inoltre, un divieto di tale portata, censurando anche eventuali contenuti parodici o satirici, avrebbe determinato una sproporzionata restrizione della libertà di manifestazione del pensiero.

Il fatto che i *deep fake* possano costituire una minaccia per l'ordine democratico, pur non potendo configurare, di per sé, un motivo valido per vietarne qualsiasi forma di riproduzione, rende necessario regolamentarne il funzionamento in modo da limitare i possibili rischi per la collettività. Sotto questo profilo, la disciplina introdotta dall'*AI Act*, sostanzialmente basata sull'imposizione di meri obblighi di informazione in capo ai *deployer*, cui non si associano misure di *enforcement* di natura sanzionatoria⁶⁵, non pare particolarmente incisiva.

Il dovere di segnalare i *deep fake* al fine di renderli facilmente riconoscibili per gli utenti può contribuire a tenere sotto controllo il fenomeno, ma difficilmente può rappresentare un deterrente utile a scongiurare eventuali utilizzi per scopi disinformativi⁶⁶. In questi casi, infatti, tali contenuti sono realizzati proprio allo scopo di manipolare gli utenti: ragion per cui gli autori avranno interesse a nascondere la natura artefatta. Com'è stato opportunamente notato altrove, se l'autore di un *deep fake* è disposto a rivelare l'origine artificiale della propria opera, è plausibile che tale operazione non celi un intento malevolo: in effetti, qualora fosse mossa da una finalità ingannevole, ad esempio poiché architettata nell'ambito di una mirata campagna di disinformazione, la diffusione di *deep fake* potrebbe essere gestita tramite strumenti informatici in grado di mascherare l'identità degli autori (come *bot* o *account* falsi), con la conseguenza di rendere piuttosto complessa l'identificazione dei soggetti in capo ai quali ricondurre la responsabilità in caso di violazione degli obblighi di notifica di cui all'art. 50, par. 4, dell'*AI Act*⁶⁷.

⁶³ In tal senso, cfr. F. R. MORENO, *Generative AI and deepfakes: a human rights approach to tackling harmful content*, in *International Review of Law, Computers & Technology*, 2024, p. 1 ss.

⁶⁴ Non è possibile, in questa sede, approfondire tutti i possibili risvolti dei *deep fake*, compresi quelli positivi. Basti ricordare che taluni studi in materia hanno evidenziato che tali tecnologie possono trovare applicazione anche per finalità creative, ad esempio in settori come il cinema, la moda e la pubblicità, nonché per scopi di sperimentazione, specialmente in ambito medico. Su questi aspetti, cfr. M. FEENEY, *Deepfake laws risk creating more problems than they solve*, Regulatory Transparency Project, 1° March 2021, reperibile *online*; J. SILBEY, W. HARTZOG, *The Upside of Deep Fakes*, in *Maryland Law Review*, 2019, p. 960 ss.

⁶⁵ In proposito, PALLADINO lamenta la mancata previsione «dell'obbligo di rimozione dei contenuti manipolati o l'introduzione di sanzioni da comminare in capo a coloro che svolgono attività di diffusione dei *deep fake* con intenti manipolatori» (R. PALLADINO, L'«*approccio europeo*» al contrasto alla disinformazione digitale e alla protezione dei valori democratici, cit., p. 315).

⁶⁶ M. LABUZ, *Regulating Deep Fakes in the Artificial Intelligence Act*, in *Applied Cybersecurity & Internet Governance*, n. 1, 2023, p. 1 ss.

⁶⁷ A. RUFFO, *Il disordine informativo e l'Intelligenza Artificiale*, cit., p. 422.

Per ovviare a questo problema, dal momento che i *deep fake* circolano prevalentemente sui *social network*, si potrebbe pensare di obbligare le piattaforme digitali a fornire indicazioni puntuali sull'origine dei contenuti pubblicati, ad esempio segnalando i casi in cui la condivisione fosse opera di un *bot*⁶⁸. Una tale previsione normativa – da concretizzare, presumibilmente, nell'ambito del DSA – pur non rimuovendo il rischio che i *deep fake* realizzati per scopi disinformativi possano sfuggire all'obbligo di notifica, consentirebbe agli utenti di interfacciarsi in modo più consapevole con i contenuti creati dall'intelligenza artificiale.

Peraltro, il regolamento non chiarisce in che modo i *deployer* dovrebbero adempiere in concreto agli obblighi di trasparenza ad essi incombenti. L'art. 50, par. 5, si limita a specificare che la segnalazione relativa alla natura manipolata di un contenuto dovrebbe avvenire «in maniera chiara e distinguibile al più tardi al momento della prima interazione o esposizione». La notifica, in altre parole, non dovrebbe lasciare spazio a dubbi, permettendo agli utenti di cogliere immediatamente l'origine artificiale di un contenuto, né essere tardiva, giacché in tal caso l'effetto manipolatorio potrebbe essersi già concretizzato, sia pur inconsciamente. Al di là di ciò, tuttavia, la disposizione lascia ai *deployer* ampia discrezionalità circa le modalità tramite cui effettuare la segnalazione. Sotto questo profilo, una maggiore uniformità sarebbe stata preferibile, così da scongiurare il rischio che gli utenti meno accorti e meno avvezzi alle tecnologie digitali possano essere indotti in errore.

Per affrontare tali carenze, il legislatore ha optato per la via della co-regolamentazione, già sperimentata nel quadro del DSA⁶⁹. L'art. 50, par. 7, affida all'Ufficio per l'IA⁷⁰ il compito di incoraggiare l'elaborazione di codici di buone pratiche, adottati dalla Commissione come atti di esecuzione ai sensi dell'art. 56, par. 6, volti a «facilitare l'efficace attuazione degli obblighi relativi alla rilevazione e all'etichettatura dei contenuti generati o manipolati artificialmente». Tali codici dovrebbero offrire, in sostanza, un supporto operativo ai destinatari degli obblighi di *compliance* stabiliti dal regolamento: tuttavia, trattandosi di strumenti di *soft law*, a cui fornitori e utilizzatori possono peraltro aderire in modo del tutto volontario, non sembrano sufficienti a colmare le lacune normative evidenziate in precedenza.

Tali criticità, infine, sono accentuate dal fatto che l'ambito di applicazione del regolamento 2024/1689 non si estenda agli utilizzi non professionali dei sistemi di IA. Ai sensi dell'art. 2, par. 10, infatti, allorché l'uso dell'intelligenza artificiale sia riconducibile a un'attività puramente personale, gli obblighi sanciti dal regolamento nei confronti dei *deployer* non trovano applicazione. Tale eccezione, benché coerente con la *ratio* dell'*AI Act*, inquadrabile in un'ottica di armonizzazione del mercato interno, solleva qualche perplessità in relazione al fenomeno dei *deep fake*⁷¹. In effetti, vista la crescente accessibilità dei *software* di intelligenza artificiale, unita alla rapidità di diffusione dei contenuti nella rete, appare difficile escludere *a priori* che un contenuto manipolato, seppur creato in un contesto prettamente privato senza apparenti finalità disinformative, possa generare conseguenze distorsive su larga scala.

⁶⁸ M. CAZZANIGA, *Una nuova tecnica (anche) per veicolare disinformazione*, cit., p. 186.

⁶⁹ Sul tema, si rinvia a E. BIRITTERI, *Contrasto alla disinformazione, Digital Services Act e attività di private enforcement*, cit.

⁷⁰ Come precisato dall'art. 3, par. 47, dell'*AI Act*, l'espressione si riferisce «alla funzione della Commissione volta a contribuire all'attuazione, al monitoraggio e alla supervisione dei sistemi di IA e dei modelli di IA per finalità generali, e della governance dell'IA prevista dalla decisione della Commissione del 24 gennaio 2024».

⁷¹ G. PROIETTI, *L'impianto regolatorio della società dell'informazione tra vecchi e nuovi equilibri*, cit., p. 347.

6. Conclusioni

L'analisi condotta ha evidenziato che l'*AI Act* affronti solo parzialmente i problemi connessi all'uso dell'intelligenza artificiale per finalità disinformative. Tale circostanza è una diretta conseguenza della scelta compiuta in proposito dal legislatore dell'Unione. Nel complesso, infatti, il regolamento si propone di promuovere lo sviluppo dei sistemi di IA, armonizzandone la progettazione, lo scambio e l'utilizzo nel mercato interno, sia pur nel rispetto dei valori sanciti dall'art. 2 TUE. Tale bilanciamento si concretizza tramite un approccio basato sul rischio che ricalca quello affermato nella legislazione in materia di sicurezza dei prodotti, dunque primariamente orientato a minimizzare gli effetti negativi dell'IA sul godimento dei diritti fondamentali. L'attenzione al dato individuale rischia, tuttavia, di lasciare in secondo piano quei processi non immediatamente lesivi di un diritto fondamentale della persona, eppure parimenti dannosi nel lungo periodo. È il caso, per l'appunto, dei fenomeni disinformativi, i quali, veicolando narrazioni manipolate, sono suscettibili non solo di alterare i meccanismi di formazione del consenso e di interferire con il regolare svolgimento dei processi elettorali, ma anche di polarizzare l'opinione pubblica e accrescere la disaffezione nei confronti della politica: sintomi evidenti di una progressiva erosione dei valori democratici.

Da questo punto di vista, la disciplina dettata dall'*AI Act* risulta piuttosto timida. Il fatto che i sistemi idonei ad influenzare l'esito delle elezioni siano inclusi, in linea di principio, tra quelli ritenuti ad alto rischio riflette la consapevolezza del legislatore circa la pericolosità dell'IA per la tenuta degli ordinamenti democratici. A tale previsione di carattere generale si accompagna, tuttavia, uno schema regolatorio che individua nella causazione di un danno significativo – di natura fisica, psicologica o economica – il presupposto necessario ai fini dell'attivazione degli obblighi di *compliance* imposti a fornitori e *deployer*. Difficilmente tale modello consentirà di contrastare adeguatamente i rischi di derive illiberali connessi all'impiego dei sistemi di IA per ragioni disinformative, giacché siffatti fenomeni producono conseguenze di natura sistemica, non contemplate nella nozione di danno accolta dal regolamento, calibrata invece sulla sfera individuale⁷².

Anche la selezione delle pratiche vietate di cui all'art. 5 del regolamento 2024/1689 non sembra offrire adeguate garanzie sul piano del contrasto alla disinformazione *online*. Come si è visto, affinché possa trovare applicazione siffatto divieto, è necessario dimostrare l'esistenza di un nesso causale diretto tra l'applicazione di IA utilizzata e l'alterazione del comportamento umano. Nel nostro caso, tale operazione rischia di rivelarsi sempre vana, dal momento che le decisioni di voto degli elettori sono estremamente volubili, nonché influenzate da una molteplicità di fattori: ne consegue che risulti difficile riscontrare, con un

⁷² Tali criticità sono solo parzialmente compensate dalla disciplina riservata ai modelli di IA a scopo generale, i quali, in ragione della molteplicità degli ambiti in cui possono trovare applicazione, possono servire un'ampia gamma di finalità e, di conseguenza, comportare rischi sistemici non prevedibili *ex ante*, tra cui rientrano, *inter alia*, gli «eventuali effetti negativi, effettivi o ragionevolmente prevedibili, sui processi democratici», nonché «la diffusione di contenuti illegali, mendaci o discriminatori». Al riguardo, l'art. 55 del regolamento *de quo* affida ai fornitori dei sistemi di IA il compito di individuare e mitigare i rischi sistemici che potrebbero derivare dai loro modelli, avvalendosi semmai dei codici di buone pratiche, strumenti di *soft law* formulati dall'Ufficio per l'IA ai sensi dell'art. 56. Tale approccio, affidandosi alla responsabilizzazione degli attori privati, non pare particolarmente efficace, rischiando, peraltro, di dare origine a forme di tutela difformi all'interno degli Stati membri. Sul punto, si rinvia a G. CONTISSA, F. GALLI, *La governance della "disinformazione aumentata" tra Digital Services Act e AI Act*, cit., p. 150 ss.

certo grado di ragionevolezza, una connessione diretta tra una campagna disinformativa avvalsa dell'intelligenza artificiale e la distorsione del comportamento elettorale delle persone. Benché comprensibile in un'ottica di certezza del diritto, tale requisito appare inidoneo ad arginare le conseguenze negative derivanti dalla disinformazione *online*, la cui natura diffusa e non misurabile non si presta ad essere ricondotta entro un rigido schema di causa-effetto.

Nel complesso, riservando maggiore attenzione all'impatto prodotto dall'intelligenza artificiale sui diritti individuali, il regolamento rischia di sottovalutarne i possibili effetti negativi sul piano collettivo – come la regressione dei principi democratici – i quali, sebbene menzionati a più riprese, non integrano sufficientemente la nozione di danno su cui, in fin dei conti, si basano sia i divieti che le limitazioni. Vero è che, a proposito del contrasto alla disinformazione, il legislatore debba intervenire con cautela, atteso che, in tale ambito, ciascun intervento restrittivo comporta una compressione della libertà di espressione e informazione. Eppure, continuare a percorrere la via europea all'evoluzione tecnologica significa, innanzitutto, agire in chiave preventiva, affinché siano neutralizzate tutte le minacce potenzialmente in grado di nuocere ai valori fondamentali di cui all'art. 2 TUE, *ivi* compresa la disinformazione *online*.



ALESSANDRO ROSANÒ*

TU CHIAMALI, SE VUOI, SISTEMI DI RICONOSCIMENTO DELLE EMOZIONI: CONSIDERAZIONI SULLA DISCIPLINA POSTA IN MATERIA DAL REGOLAMENTO (UE) 2024/1689

SOMMARIO: 1. Introduzione. – 2. I sistemi di riconoscimento delle emozioni: una breve panoramica su concetti, applicazioni e progetti finanziati dall'Unione europea. – 3. I sistemi di riconoscimento delle emozioni nell'*AI Act*: dalla proposta della Commissione europea al regolamento (UE) 2024/1689. – 4. Profili critici: la protezione dei dati personali. – 5. Profili critici: le pratiche vietate. – 6. Conclusioni.

1. Introduzione

Che cosa si intenda per emozione è tema dibattuto, tanto che, nel corso del tempo, è stato proposto un numero particolarmente alto – tendente al centinaio – di definizioni¹ e che ancora oggi gli studiosi hanno difficoltà a concordare quanto a un concetto condiviso².

In linea assolutamente generale, può prendersi per buona la nozione secondo la quale si tratterebbe di una condizione cognitiva e fisiologica dinamica che si sviluppa in reazione a *input* come esperienze, pensieri o interazioni con altre persone³. Dunque, sulla base di questo, può sostenersi che l'esistenza dell'essere umano è fortemente contraddistinta dalle emozioni e che la comprensione di esse è centrale affinché il singolo possa non solamente assumere una migliore consapevolezza di sé, ma altresì relazionarsi appieno con altri soggetti.

Al riguardo, un'idea è stata fatta oggetto di ampia riflessione almeno dalla seconda metà del secolo scorso: quella secondo cui sarebbe possibile individuare alcune emozioni umane

* Ricercatore *Tenure Track* in Diritto dell'Unione europea, Università degli Studi di Parma.

¹ Per una panoramica, ancorché non aggiornata, P.R. KLEINGINNA, A.M. KLEINGINNA, *A Categorized List of Emotion Definitions, with Suggestions for a Consensual Definition*, in *Motivation and Emotion*, 1981, p. 345 ss.

² C.E. IZARD, *The Many Meanings/Aspects of Emotion: Definitions, Functions, Activation, and Regulation*, in *Emotion Review*, 2010, p. 367. Per ulteriori riferimenti, senza pretesa di esaustività, S.C. WIDEN, J.A. RUSSELL, *Descriptive and Prescriptive Definitions of Emotion*, in *Emotion Review*, 2010, p. 377 ss. e K. MULLIGAN, K.R. SCHERER, *Toward a Working Definition of Emotion*, in *Emotion Review*, 2012, p. 345 ss.

³ S.K. KHARE ET AL., *Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations*, in *Information Fusion*, 2024, p. 102019. Per una più ampia ricostruzione, R.A. CALVO ET AL. (eds), *The Oxford Handbook of Affective Computing*, Oxford, 2014.

di base, ossia componenti psicologici elementari che permettono di comprendere e spiegare il ventaglio di emozioni che l'individuo è in grado di provare. A questo si ricollega la convinzione che rappresenta il fondamento di una branca dell'intelligenza artificiale (nel prosieguo, IA), l'*Affective Computing*, nell'ambito della quale si intende studiare e predisporre metodi attraverso cui i computer possano identificare, processare, comprendere e replicare le emozioni umane. Dunque, emerge tra l'altro la questione dello sviluppo e dell'utilizzo di sistemi di IA in grado di riconoscere le emozioni.

Non si tratta di fantascienza. Nel 2007, all'Università di Oxford, fu fondata *Realeyes*, che oggi si presenta come «the global leader in AI that understands human behavior. We use webcams to understand the identity, attention and emotions of users who opt-in to ultimately make digital experiences better – more human»⁴. Nel 2009, in Massachusetts, nel contesto del MIT, fu costituita *Affectiva*, una società di sviluppo di software di IA specializzati nella rilevazione di emozioni e stati cognitivi a partire dall'analisi del volto, della voce e dei movimenti del corpo degli esseri umani⁵.

Per quanto i sistemi in parola sembrano certamente delineare prospettive interessanti a livello imprenditoriale, è altresì indubbio che sollevino interrogativi importanti per il giurista. Infatti, essi paiono confermare la convinzione che si sia passati da «tecnologie che un tempo operavano principalmente nel mondo virtuale a sistemi computazionali e algoritmici che sfruttando, invece, elementi del mondo reale, invadono la sfera personale degli utenti e mirano a modificare le azioni nel mondo reale»⁶. Si tratta di quella che Shoshana Zubioff ne *Il capitalismo della sorveglianza* ha definito come «la nuova frontiera del *surplus* comportamentale, nella quale il continente dei *dark data* della nostra vita (...) viene portato alla luce perché qualcun altro possa guadagnarci»⁷. Per la verità, non di solo guadagno economico può parlarsi, visto che pare che nella Repubblica popolare cinese sistemi di riconoscimento delle emozioni sarebbero utilizzati per sorvegliare la minoranza degli Uiguri⁸. Quindi, si pone un problema di tutela dei diritti fondamentali dell'individuo sia sul piano del rapporto con i nuovi potentati economici dell'ambito digitale, sia su quello, più tradizionale, del rapporto con il potere pubblico. Proprio questo genera l'esigenza di intervenire normativamente, al fine di regolare l'utilizzo dei sistemi qui in discussione⁹.

Allora, nel paper, chiariti alcuni aspetti tecnici relativi al riconoscimento delle emozioni, ai sistemi sviluppati a tal fine e ai loro utilizzi (paragrafo 2) si intende illustrare la disciplina posta dal regolamento (UE) 2024/1689 (il cosiddetto *AI Act*), mettendo in rilievo come le disposizioni in materia siano molto cambiate nel passaggio dalla proposta della Commissione europea al testo definitivo (paragrafo 3). In seguito, si prendono in considerazione le questioni sollevate dal ricorso a tali sistemi rispetto alla protezione dei dati personali (paragrafo 4) e ci si concentra sulle difficoltà ermeneutiche poste dal divieto quanto ad alcune pratiche ai sensi dell'art. 5 del regolamento (paragrafo 5). Le conclusioni riassumono l'analisi svolta in questa sede (paragrafo 6).

⁴ <https://www.realeyesit.com/company/about-us/>.

⁵ <https://www.affectiva.com/about-affectiva/>.

⁶ E.M. INCUTTI, *Sistemi di riconoscimento delle emozioni e ruolo dell'autonomia privata: linee evolutive di un umanesimo digitale*, in *Giustizia civile*, 2022, p. 521, nota 13.

⁷ S. ZUBOFF, *Il capitalismo della sorveglianza*, Roma, 2019, p. 269 (traduzione italiana).

⁸ J. WAKEFIELD, *AI emotion-detection software tested on Uyghurs*, in *BBC.com*, 26 maggio 2021, disponibile online.

⁹ K. CRAWFORD, *Time to Regulate AI that Interprets Human Emotions*, in *Nature*, 6 aprile 2021, disponibile online.

2. I sistemi di riconoscimento delle emozioni: una breve panoramica su concetti, applicazioni e progetti finanziati dall'Unione europea

Una persona, quando reagisce a un determinato evento, trasmette un'emozione. Secondo certe teorie, esistono alcune emozioni di base generalmente individuabili in tutte le culture¹⁰. Al riguardo, vi è stato chi le ha identificate in paura, rabbia, angoscia, gioia, disgusto, sorpresa, interesse e vergogna¹¹, chi ha fatto riferimento a paura, rabbia, dolore, gioia, disgusto, sorpresa, accettazione e attesa¹² e chi, dopo avere scorso in rassegna varie proposte classificatorie, ha concluso che le quattro emozioni più ricorrenti sono paura, rabbia, tristezza e gioia, seguite da disgusto e sorpresa, mentre, quanto a tutte le altre, si assiste a divergenze significative a seconda degli autori presi in considerazione¹³.

A prescindere dall'impostazione adottata, è innegabile che le emozioni possano essere espresse attraverso sia il linguaggio parlato, sia altre forme di comunicazione, quali movimenti del corpo e tono della voce¹⁴, ossia mediante modalità verbali e non verbali che possono costituire l'oggetto di un'attività interpretativa, vale a dire di un'attività funzionale a ottenere informazioni circa lo stato emotivo di un soggetto.

Il dato interessante risiede nel fatto che la lettura di quelle reazioni può essere svolta tanto da un altro essere umano, quanto da un sistema di IA. Si tratta qui di considerare lo sviluppo dell'*Affective Computing*, un'area di ricerca multidisciplinare, dato che, per conseguire le proprie finalità, si avvale di contributi provenienti da psicologia, fisiologia, ingegneria, sociologia, matematica, informatica, scienze dell'educazione e linguistica. Ciò dipende dalla complessità degli obiettivi perseguiti dall'*Affective Computing*, i quali riguardano la creazione di e l'interazione con macchine capaci di riconoscere le emozioni dell'utente (riconoscimento emotivo), influenzarne lo stato emotivo (manipolazione emotiva), riprodurre (espressione emotiva) o provare emozioni (sintesi emotiva)¹⁵. Dunque, i temi di fondo che vengono affrontati nel contesto dell'*Affective Computing* sono due: il riconoscimento delle emozioni e l'analisi delle stesse¹⁶. Si parla di *Emotion AI* per fare riferimento a quei sistemi che analizzano le espressioni facciali e altre manifestazioni fisiche delle emozioni e che creano risposte emotive¹⁷.

La tecnologia alla base dell'*Emotion AI* si fonda su metodi generalizzabili che permetterebbero di identificare le emozioni altrui con modalità paragonabili a quelle a disposizione di un essere umano, se non addirittura migliori¹⁸. In particolare, hanno assunto sempre più rilievo quelli che si basano sull'analisi delle espressioni facciali, del tono della

¹⁰ M.N. SHIOTA, *Basic and Discrete Emotion Theories*, in A. SCARANTINO (ed.), *Emotion Theory: The Routledge Comprehensive Guide. Volume I: History, Contemporary Theories, and Key Elements*, New York, 2024, p. 310 ss.

¹¹ S.S. TOMKINS, *Affect, imagery, consciousness: Vol. I. The positive affect*, New York, 1962.

¹² R. PLUTCHIK, *Emotion: A Psychoevolutionary Synthesis*, New York, 1980.

¹³ A. ORTONY, G.L. CLORE, A. COLLINS, *The cognitive structure of emotions*, Cambridge, 1988.

¹⁴ A.A. VARGHESE, J.P. CHERIAN, J.J. KIZHAKKETHOTTAM, *Overview on emotion recognition system*, in *2015 International Conference on Soft-Computing and Networks Security (ICSNS)*, 2015, p. 1.

¹⁵ Per un'introduzione, R.W. PICARD, *Affective Computing*, Cambridge (MA), 1997 e S.B. DAILY ET AL., *Affective Computing: Historical Foundations, Current Applications, and Future Trends*, in M. JEON (ed.), *Emotions and affect in human factors and human-computer interaction*, Cambridge (MA), 2017, p. 213 ss.

¹⁶ Y. WANG ET AL., *A systematic review on affective computing: emotion models, databases, and recent advances*, in *Information Fusion*, 2022, p. 19.

¹⁷ M. SOMERS, *Emotion AI, Explained*, in MIT Sloan, 8 marzo 2019, disponibile online.

¹⁸ J.S. BARD, *Developing a Legal Framework for Regulating Emotion AI*, in *Boston University Journal of Science and Technology Law*, 2021, p. 279.

voce, di segnali di natura fisiologica (ricorrendo per esempio all'elettrocardiografia o all'elettroencefalografia o alla risonanza magnetica funzionale) o di testi scritti¹⁹. A seconda del fatto che ricorra a una soltanto di queste modalità o a una combinazione delle stesse, si distingue tra sistemi di riconoscimento unimodali e multimodali. Dato che l'approccio unimodale non sembra abbastanza affidabile, la tendenza crescente nell'ambito dell'*Affective Computing* è nel senso di ricorrere a forme multimodali di riconoscimento delle emozioni²⁰. L'accuratezza può variare, anche se tende a superare il 90%²¹.

Le applicazioni dei sistemi di riconoscimento delle emozioni sono molteplici. Nel settore sanitario, essi si rivelano utili per quel che riguarda la rilevazione di condizioni fisiologiche quali affaticamento²² e dolore²³, nonché l'identificazione e la cura di vari disturbi di carattere neurologico, come la malattia di Parkinson²⁴, i disturbi del sonno²⁵, la schizofrenia e la depressione²⁶. Con riferimento a quest'ultimo aspetto, si può ricordare il progetto *Predicting Response to Depression Treatment (PReDicT)*, che ha ottenuto poco più di quattro milioni di euro da parte di Horizon 2020²⁷. Alla base del progetto stava l'esigenza di arrivare a comprendere rapidamente se un trattamento antidepressivo stia funzionando o meno. Normalmente, perché un medico possa fare questo, sono necessarie dalle quattro alle sei settimane e, nella maggior parte dei casi, i pazienti non rispondono al primo trattamento, con la conseguenza che possono servire molti mesi prima che si arrivi a identificare la cura corretta. Pertanto, nell'ambito del progetto, si è cercato di sviluppare un test, fondato sul rilevamento della risposta emotiva dei pazienti, che avrebbe permesso di stabilire entro una settimana dall'inizio del trattamento se un antidepressivo stesse facendo effetto, così da contenere i tempi richiesti per individuare la cura e i costi sanitari ed economico-sociali della depressione. Nell'85% dei casi, i pazienti hanno mantenuto il trattamento indicato per primo sulla base della valutazione condotta grazie al test²⁸.

Inoltre, i governi e le organizzazioni internazionali possono ricorrere ai sistemi di riconoscimento delle emozioni per svolgere funzioni connesse all'esercizio dei loro poteri, come per quel che riguarda il controllo dei confini. Il caso probabilmente più noto è quello

¹⁹ Per un'introduzione, si rinvia a J.M. HARLEY, *Measuring Emotions: A Survey of Cutting Edge Methodologies Used in Computer-Based Learning Environment Research*, in S.Y. TETTEGAH, M. GAARTMEIER (eds), *Emotions, Technology, Design, and Learning*, Cambridge, 2016, p. 89 ss. e A. SAXENA, A. KHANNA, D. GUPTA, *Emotion Recognition and Detection Methods: A Comprehensive Survey*, in *Journal of Artificial Intelligence and Systems*, 2020, p. 53 ss.

²⁰ Al riguardo, G. CARIDAKIS ET AL., *Multimodal emotion recognition from expressive faces, body gestures and speech*, in *Artificial intelligence and innovations 2007: From theory to applications*, Berlin, 2007, p. 375 ss. e T. VERHOEF, E. FOSCH-VILLARONGA, *Towards affective computing that works for everyone*, in *2023 11th International Conference on Affective Computing and Intelligent Interaction*, p. 1 ss.

²¹ S.K. KHARE ET AL., cit.

²² J. SUN ET AL., *Memristor-based neural network circuit of emotion congruent memory with mental fatigue and emotion inhibition*, in *IEEE Transactions on Biomedical Circuits and Systems*, 2021, p. 606 ss.

²³ P. LUCEY ET AL., *Automatically detecting pain in video through facial action units*, *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 2011, p. 664 ss.

²⁴ M.N. DAR ET AL., *EEG-based emotion charting for Parkinson's disease patients using Convolutional Recurrent Neural Networks and cross dataset learning*, in *Computers in Biology and Medicine*, 2022, p. 105327 ss.

²⁵ R.E. DAHL, A.G. HARVEY, *Sleep in children and adolescents with behavioral and emotional disorders*, in *Sleep Medicine Clinics*, 2007, p. 501 ss.

²⁶ T.E. FEINBERG ET AL., *Facial discrimination and emotional recognition in schizophrenia and affective disorders*, in *Archives of General Psychiatry*, 1986, p. 276 ss.

²⁷ <https://cordis.europa.eu/project/id/696802>.

²⁸ Per un approfondimento, M. BROWNING ET AL., *The clinical effectiveness of using a predictive algorithm to guide antidepressant treatment in primary care (PReDicT): an open-label, randomised controlled trial*, in *Neuropsychopharmacology*, 2021, p. 1307 ss.

del progetto *Intelligent Portable Border Control System*, meglio noto con l'acronimo *iBorderCtrl*, finanziato da Horizon 2020 con poco più di quattro milioni e mezzo di euro e realizzato tra il 2016 e il 2019, il quale mirava a introdurre controlli di frontiera più rapidi e approfonditi, in quanto basati su sistemi tecnologicamente avanzati, nei confronti di cittadini di Stati terzi che avessero attraversato i confini dell'Unione²⁹. Più precisamente, si prevedeva il ricorso a una procedura articolata, nell'ambito della quale, tra l'altro, i cittadini degli Stati terzi sarebbero stati sottoposti a un'intervista con un *avatar* al fine di rilevare eventuali menzogne. Pertanto, ai valichi di frontiera di Grecia, Lettonia e Ungheria, i cittadini di Stati terzi hanno dovuto rispondere ad alcune domande poste dal *silent talker*, un sistema di analisi in grado di registrare e analizzare i cosiddetti biomarcatori dell'inganno, ossia micromovimenti del volto, microespressioni facciali non verbali (come battiti di ciglia, arrossamento e movimenti della testa), così da comprendere se il soggetto intervistato fosse sincero o meno. Nel primo caso, è stato permesso l'attraversamento delle frontiere; nel secondo, si è reso necessario fornire informazioni aggiuntive. Il tasso di accuratezza è stato del 73-75%³⁰.

I sistemi di riconoscimento delle emozioni possono altresì essere integrati in prodotti di consumo. Si pensi al caso delle automobili che, così, possono registrare cosa accade all'interno del veicolo e rilevare emozioni e stati cognitivi in modo da porvi rimedio (come nell'ipotesi in cui il conducente sia colto da sonnolenza o distrazione)³¹.

Ulteriormente, i sistemi di riconoscimento delle emozioni possono essere utilizzati nell'ambito: dei colloqui di lavoro, per monitorare il livello di stress dei candidati; dell'istruzione, per comprendere quale sia il livello di soddisfazione degli studenti e arrivare a sviluppare contenuti *ad hoc* a seconda delle diverse capacità di ciascuno; del marketing, per analizzare in tempo reale le reazioni dei clienti e migliorare le esperienze di questi nei negozi³². In relazione agli ultimi due settori ora menzionati, si possono ricordare due progetti finanziati da Horizon 2020: *DE-ENIGMA: Multi-Modal Human-Robot Interaction for Teaching and Expanding Social Imagination in Autistic Children*, inteso a migliorare i sistemi di riconoscimento delle emozioni utilizzati nelle terapie psicoeducative dirette a bambini autistici³³; *Social Semantic Emotion Analysis for Innovative Multilingual Big Data Analytics Markets (MixedEmotions)*, attraverso il quale si è cercato di realizzare applicazioni capaci di svolgere un'analisi multimodale delle emozioni degli utenti così da permettere di affrontare questioni connesse alla gestione della reputazione dei marchi e allo svolgimento delle operazioni dei *call center*³⁴.

²⁹ <https://cordis.europa.eu/project/id/700626>.

³⁰ Per un approfondimento, J. SÁNCHEZ-MONEDERO, L. DENCİK, *The politics of deceptive borders: 'biomarkers of deceit' and the case of iBorderCtrl*, in *Information, Communication & Society*, 2022, p. 413 ss.

³¹ G. DERY, *Is a Friend Truly a Friend If You Can Just Leave it In the Garage? Toyota's and Honda's Concept Cars Could Have Significant Fourth Amendment Implications*, in *American Criminal Law Review*, 2018, p. 585 ss.

³² Per questi e ulteriori esempi, L. CEN ET AL., *A real-time speech emotion recognition system and its application in online learning*, in S.Y. TETTEGAH, M. GARTMEIER (eds), cit., p. 27 ss. e D. CARUELLE ET AL., *Affective computing in marketing: practical implications and research opportunities afforded by emotionally intelligent machines*, in *Marketing Letters*, 2022, p. 163 ss.

³³ <https://cordis.europa.eu/project/id/688835>.

³⁴ <https://cordis.europa.eu/project/id/644632>.

3. *I sistemi di riconoscimento delle emozioni nell'AI Act: dalla proposta della Commissione europea al regolamento (UE) 2024/1689*

Per quel che riguarda i sistemi di riconoscimento delle emozioni, va detto che la normativa rilevante, posta ora dall'*AI Act*, è cambiata molto nel passaggio dalla proposta della Commissione al testo definitivo. Nella proposta si prevedeva che il regolamento avrebbe stabilito, *inter alia*, regole di trasparenza armonizzate per i sistemi di IA, tra i quali i sistemi di riconoscimento delle emozioni (art. 1, par. 1, lett. c)). Questi ultimi erano definiti come sistemi di IA finalizzati all'identificazione o alla deduzione di emozioni o intenzioni di persone fisiche sulla base dei loro dati biometrici, ossia di dati personali ottenuti da un trattamento tecnico specifico relativi alle caratteristiche fisiche, fisiologiche o comportamentali di una persona fisica che ne avrebbero consentito l'identificazione univoca, come nel caso dell'immagine facciale o dei dati dattiloscopici (art. 3, n. 33) e 34)). All'allegato III si chiariva che rientravano tra i sistemi ad alto rischio quelli destinati a essere utilizzati dalle autorità di contrasto o dalle autorità pubbliche competenti in materia di gestione della migrazione, dell'asilo e del controllo delle frontiere, come poligrafi e strumenti analoghi, o per rilevare lo stato emotivo di una persona fisica. Di conseguenza, tutti gli altri sistemi di riconoscimento delle emozioni sarebbero rientrati tra quelli a basso rischio, in relazione ai quali avrebbero trovato applicazione l'obbligo di informare le persone fisiche quanto al fatto di stare interagendo con un sistema di IA, a meno che ciò non risultasse evidente dalle circostanze e dal contesto di utilizzo, e quanto al funzionamento del sistema (art. 52, par. 1 e 2).

Da parte sua, il Consiglio dell'Unione europea fece, nel proprio orientamento generale sulla proposta, un numero limitato di interventi, alcuni dei quali risultavano però di non poco rilievo. Soprattutto, in relazione ai sistemi classificati come ad alto rischio nell'allegato III, si indicava che tali avrebbero dovuto essere non solamente quelli destinati a essere utilizzati dalle autorità di contrasto o dalle autorità pubbliche competenti in materia di gestione della migrazione, dell'asilo e del controllo delle frontiere, come poligrafi e strumenti analoghi, o per rilevare lo stato emotivo di una persona fisica, ma altresì i sistemi di tale tipo utilizzati per conto delle suddette autorità. Inoltre, quanto agli obblighi di trasparenza, si specificava che l'obbligo di informazione non avrebbe trovato applicazione per i sistemi di IA per il riconoscimento delle emozioni autorizzati dalla legge per accertare, prevenire e indagare reati, fatte salve le tutele adeguate per i diritti e le libertà dei terzi (art. 52, par. 2 *bis*)³⁵.

Il Parlamento europeo fu estremamente critico, esprimendo dubbi sulla base scientifica dei sistemi di riconoscimento delle emozioni in ragione del fatto che queste ultime, la loro espressione e la loro percezione sono influenzate da fattori di ordine culturale e dalle singole situazioni, tendendo peraltro a cambiare da soggetto a soggetto³⁶. Perciò, il Parlamento europeo riteneva necessario inserire tra le pratiche vietate l'immissione sul mercato, la messa in servizio o l'uso di sistemi di riconoscimento delle emozioni destinati a essere utilizzati

³⁵ Proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione – Orientamento generale (6 dicembre 2022), 15698/22.

³⁶ Emendamenti del Parlamento europeo, approvati il 14 giugno 2023, alla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), emendamento 52, considerando 26 *quater* (nuovo).

nell'ambito della gestione delle frontiere, per attività di contrasto, sul luogo di lavoro e negli istituti di istruzione e classificare come ad alto rischio tutti gli altri sistemi di riconoscimento delle emozioni³⁷.

La disciplina oggi in vigore è quella risultante dall'*AI Act*, ai sensi del quale i sistemi di riconoscimento delle emozioni sono finalizzati all'identificazione o all'inferenza di emozioni o intenzioni di persone fisiche sulla base dei loro dati biometrici (art. 3, n. 39). Quindi, essi si concentrano su profili quali felicità, tristezza, rabbia, sorpresa, disgusto, imbarazzo, eccitazione, vergogna, disprezzo, soddisfazione e divertimento e non prendono in considerazione stati fisici, come dolore o affaticamento, o semplici espressioni, gesti o movimenti immediatamente evidenti, a meno che non vengano utilizzati per identificare o inferire emozioni (considerando 18). Posto che sono stati avanzati dubbi quanto alla base scientifica di tali sistemi e che la loro limitata affidabilità, mancanza di specificità e limitata generalizzabilità possono condurre a risultati discriminatori o comunque a esiti lesivi dei diritti fondamentali, soprattutto in contesti lavorativi ed educativi (considerando 44), rientrano tra le pratiche vietate l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di sistemi di IA per inferire le emozioni di una persona fisica nell'ambito del luogo di lavoro e degli istituti di istruzione, tranne laddove i suddetti sistemi siano destinati a essere messi in funzione o immessi sul mercato per motivi medici o di sicurezza (art. 5, par. 1, lett. f). Tutti gli altri sistemi di riconoscimento delle emozioni, invece, vengono qualificati come ad alto rischio (allegato III).

4. *Profili critici: la protezione dei dati personali*

Come è stato notato, per quanto le emozioni si configurino senza dubbio come qualcosa di strettamente personale, perché legate ai valori di una persona e a ciò a cui quella persona tiene, esse non possono essere considerate di per se stesse dati personali³⁸. Risaputamente, per dato personale si intende qualsiasi informazione riguardante una persona fisica identificata o identificabile ed è identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un elemento quale il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo *online* o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale (art. 4, n. 1) del regolamento (UE) 2016/679 sulla protezione dei dati, meglio noto con l'acronimo inglese GDPR). Allora, è evidente che il diritto dell'Unione non può assicurare tutela ai dati emotivi in quanto tali e si comprende perché la definizione di sistema di riconoscimento delle emozioni posta dall'*AI Act* si concentra sul fatto che le emozioni e le intenzioni possano essere dedotte a partire dai dati biometrici di una persona, ossia da dati personali ottenuti da un trattamento tecnico specifico relativi alle caratteristiche fisiche, fisiologiche o comportamentali di un soggetto, come le immagini facciali o i dati dattiloscopici (art. 3, n. 34) e n. 39)). Ne consegue che le garanzie previste dal GDPR relative a liceità, correttezza e trasparenza del trattamento, limitazione delle finalità,

³⁷ *Ini*, emendamento 226, Articolo 5 – paragrafo 1 – lettera d) *quater* (nuova), emendamento 63, Considerando 33 *bis* (nuovo) ed emendamento 712, Allegato III – punto 1 – lettera a) *bis* (nuova).

³⁸ A. HÄUSELMANN, *Fit for purpose? Affective computing meets EU data protection law*, in *International Data Privacy Law*, 2021, p. 245 ss. e A. HÄUSELMANN ET AL., *EU law and emotion data*, in *11th International Conference on Affective Computing and Intelligent Interaction*, 2023, p. 3.

minimizzazione dei dati, esattezza, limitazione della conservazione, integrità e riservatezza devono trovare applicazione anche per quel che riguarda i dati biometrici ottenuti attraverso i sistemi di riconoscimento delle emozioni.

Ci si può chiedere, poi, se possa ritenersi che i dati in parola rientrino tra i dati sensibili (art. 9 GDPR), il trattamento dei quali è vietato, tranne che in alcuni casi. Nella categoria in questione rientrano altresì i dati biometrici «intesi a identificare in modo univoco una persona fisica». Si tratta, come è stato detto, di una fattispecie *purpose-based*, attraverso la quale si mira a evitare un'automatica sovrapposizione tra dati biometrici e dati sensibili quando i primi siano stati ottenuti senza che vi fosse effettivamente l'intenzione di procedere a un'identificazione³⁹. Perciò, affinché i dati biometrici possano essere considerati dati sensibili, bisogna stabilire quale sia la finalità del trattamento, ossia se i dati siano stati ottenuti per identificare il soggetto o meno. Ora, è vero che i sistemi di riconoscimento delle emozioni si basano sull'analisi delle espressioni facciali, del tono della voce e di segnali di natura fisiologica risultanti da elettrocardiografia, elettroencefalografia o risonanza magnetica funzionale, che rientrano nella nozione di dato biometrico, ma solamente nel caso concreto si potrà stabilire se tali dati biometrici siano stati acquisiti per identificare in modo univoco una persona.

Resta il fatto, comunque, che le informazioni risultanti dall'elettrocardiografia, dall'elettroencefalografia e dalla risonanza magnetica funzionale sicuramente si configurano come dati relativi alla salute, vale a dire dati personali attinenti alla salute fisica o mentale di una persona fisica, compresa la prestazione di servizi di assistenza sanitaria, che rivelano informazioni relative al suo stato di salute (art. 4, n. 15), GDPR) e, di conseguenza, come dati sensibili. A conferma di ciò, si consideri che, secondo quanto suggerito nel preambolo del GDPR, tra i dati relativi alla salute dovrebbero rientrare anche le informazioni risultanti da esami e controlli effettuati su una parte del corpo e qualsiasi informazione riguardante trattamenti clinici o lo stato fisiologico o biomedico dell'interessato (considerando 35)⁴⁰.

Relativamente all'ipotesi in cui il sistema di riconoscimento delle emozioni considerato operi attraverso l'analisi di testi scritti, premesso che a una simile tecnologia non rientra tra quelle contemplate dall'*AI Act* quanto ai sistemi di riconoscimento delle emozioni perché non fa ricorso a dati biometrici, è chiaro che eventuali questioni che possano sorgere non attengono alla protezione dei dati personali ma, eventualmente, alla tutela della *privacy*.

5. *Profili critici: le pratiche vietate*

Come detto in precedenza, tra le pratiche vietate dall'*AI Act* rientrano l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di sistemi di IA per inferire le emozioni di una persona fisica nell'ambito del luogo di lavoro e degli istituti di istruzione, tranne laddove l'uso del sistema di IA sia destinato a essere messo in funzione o immesso sul mercato per motivi medici o di sicurezza, come i sistemi destinati all'uso terapeutico (art.

³⁹ In questi termini, P. QUINN, G. MALGIERI, *The Difficulty of Defining Sensitive Data – The Concept of Sensitive Data in the EU Data Protection Framework*, in *German Law Journal*, 2021, p. 1594. Si veda inoltre C. JASSERAND, *Legal Nature of Biometric Data: From 'Generic' Personal Data to Sensitive Data*, in *European Data Protection Law Review*, 2016, p. 297 ss.

⁴⁰ *Mutatis mutandis*, si riprendono considerazioni svolte in A. ROSANÒ, *Tra Neuromanti e pecore elettriche: Metaverso e tutela dei dati personali e della privacy*, in *Media Laws*, 2025 (in corso di pubblicazione).

5, par. 1, lett. f) e considerando 44). La scelta di limitare il divieto ai casi dei sistemi usati nei contesti lavorativi e negli istituti di istruzione è stata ricollegata allo squilibrio di poteri che in essi si può verificare (considerando 44). Come confermato dalla Commissione europea nelle linee guida sulle pratiche vietate⁴¹, l'espressione «luogo di lavoro» va intesa in senso ampio, ricomprendendo qualsiasi spazio fisico o virtuale in cui le persone fisiche si impegnano in compiti e responsabilità assegnati dal loro datore di lavoro o dall'organizzazione di cui fanno parte e assicurando tutela anche ai candidati durante il processo di selezione e assunzione⁴². Pure il riferimento ai «luoghi di istruzione» va fatto oggetto di interpretazione estensiva, considerando istituzioni pubbliche e private, senza alcuna limitazione connessa all'età degli allievi o all'ambiente specifico in cui si svolgono le attività di insegnamento⁴³.

La formulazione del divieto implica che, nell'ipotesi in cui i sistemi di riconoscimento siano utilizzati nell'ambito del luogo di lavoro e degli istituti di istruzione per motivi medici o di sicurezza, essi si configurano come sistemi ad alto rischio⁴⁴. Al riguardo, la Commissione ha sottolineato che l'eccezione relativa ai suddetti motivi va interpretata restrittivamente⁴⁵, aggiungendo che «therapeutic uses should be understood to mean uses of CE-marked medical devices. Moreover, this exception does not comprise the use of emotion recognition systems to detect general aspects of wellbeing. The general monitoring of stress levels at the workplace is not permitted under health or safety aspects (...). The notion of safety reasons within this exception should be understood to apply only in relation to the protection of life and health and not to protect other interests, for example property against theft or fraud. It follows from this narrow interpretation of the exception that any use for medical and safety reasons should always remain limited to what is strictly necessary and proportionate, including limits in time, personal application and scale, and should be accompanied by sufficient safeguards»⁴⁶.

In generale e in un'ottica di rafforzamento della tutela dei diritti fondamentali, non può che convenirsi con la Commissione europea. Tuttavia, vi sono alcune ragioni per ritenere che la posizione espressa nelle linee guida (che, comunque, non sono vincolanti) possa essere contestata.

Infatti, secondo la Commissione, «this exception does not comprise the use of emotion recognition systems to detect general aspects of wellbeing». Eppure, sempre nelle linee guida, uno degli esempi proposti è il seguente: «An employer would be prohibited from deploying AI-enabled devices or digital assistants at the workplace for measuring anxiety based on measured stress levels or for measuring boredom of employees, unless the elevated stress level/lack of concentration would pose a specific danger»⁴⁷. Sembra quindi potersi registrare una tensione tra l'affermazione astratta quanto al fatto che l'eccezione non assume rilievo

⁴¹ Coerentemente a quanto disposto dall'art. 96, par. 1, lett. b), del regolamento (UE) 2024/1689, la Commissione europea ha elaborato orientamenti sull'attuazione pratica dell'art. 5, pubblicati il 4 febbraio 2025 (Communication to the Commission, *Approval of the content of the draft Communication from the Commission - Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)*, C(2025) 884 final).

⁴² Annex to the Communication to the Commission, *Approval of the content of the draft Communication from the Commission - Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)*, C(2025) 884 final, punto 254.

⁴³ *Ivi*, punto 255.

⁴⁴ *Ivi*, punto 37.

⁴⁵ *Ivi*, punto 256.

⁴⁶ *Ivi*, punti 257-259.

⁴⁷ *Ivi*, punto 257.

con riferimento a questioni concernenti il benessere di una persona e situazioni concrete che invece potrebbero giustificare ciò. Del resto, va ricordato che dalla giurisprudenza della Corte di giustizia⁴⁸, confermata anche dal Comitato europeo per la protezione dei dati⁴⁹, risulta che la nozione di dati relativi alla salute di una persona deve costituire oggetto di un'interpretazione ampia, tale da ricomprendere tutti gli aspetti inerenti alla salute fisica e psichica di un soggetto, e quindi non contenendo il riferimento a situazioni patologiche, ma prendendo in considerazione la fisiologia dell'essere umano nel suo complesso. Dunque, nel diritto dell'Unione, il concetto di salute va interpretato come sinonimo di benessere psicofisico e i motivi medici andrebbero ricollegati a ciò.

Comunque, anche volendo ritenere che tali motivi possano essere interpretati restrittivamente, è la stessa Commissione a richiamare, per quanto riguarda i motivi di sicurezza, la protezione della salute, che deve essere necessariamente intesa in senso ampio.

Inoltre, circa i motivi di sicurezza, c'è un tema ampio che merita di essere quanto meno accennato e che attiene a come debba essere intesa la nozione stessa di sicurezza. L'argomento è stato e costituisce ancora oggetto di discussione⁵⁰, atteso che ragioni di sicurezza pubblica possono giustificare limitazioni alla libertà di circolazione delle merci (art. 36 TFUE), dei lavoratori (art. 45 TFUE), dei servizi (art. 52 TFUE) e dei capitali (art. 65 TFUE), che ogni Stato membro può adottare le misure che ritenga necessarie alla tutela degli interessi essenziali della propria sicurezza (art. 346 TFUE) e che l'Unione è tenuta a rispettare le funzioni essenziali degli Stati membri, compresa quella di tutela della sicurezza nazionale (art. 4, par. 2, TUE).

La questione ha dunque portata trans-settoriale, tanto che si può affermare che difficilmente possono essere identificati ambiti non toccati dal tema della sicurezza⁵¹, e la sua comprensione è resa ancor più difficile dall'evoluzione del contesto socio-politico⁵², soprattutto in ragione degli sviluppi di natura tecnologica⁵³.

Volendo provare a definire il concetto da un punto di vista pubblicistico, può prendersi per buono quanto sostenuto dalla Corte di giustizia, secondo la quale la competenza degli Stati membri in materia corrisponde all'interesse primario di tutelare le loro funzioni essenziali e gli interessi fondamentali della società e comprende la prevenzione e la repressione di attività tali da destabilizzare gravemente le strutture costituzionali, politiche, economiche o sociali fondamentali di un Paese e, soprattutto, da minacciare direttamente la società, la popolazione o gli Stati in quanto tali, quali in particolare le attività di terrorismo⁵⁴. Pertanto, potrebbe ritenersi che sia alla luce di tale nozione che vadano intesi i motivi di sicurezza sopra menzionati.

⁴⁸ Corte giust., 6 novembre 2003, C-101/01, *Lindqvist*, punto 50, ECLI:EU:C:2003:596 e Trib., 3 dicembre 2015, T-343/13, CN / *Parlamento europeo*, punto 50, ECLI:EU:T:2015:926.

⁴⁹ Comitato europeo per la protezione dei dati, *Linee-guida 03/2020 sul trattamento dei dati relativi alla salute a fini di ricerca scientifica nel contesto dell'emergenza legata al COVID-19*, 21 aprile 2020, punto 7.

⁵⁰ Per un'introduzione al tema e i necessari riferimenti, S. PUGLIESE, *L'evoluzione della nozione di sicurezza dell'Unione europea tra crisi sistemiche, innovazione tecnologica e tutela dei valori*, Bari, 2025.

⁵¹ Editorial, *The Passion for Security in European Societies*, in *Common Market Law Review*, 2024, p. 283.

⁵² S. BARBOU DES PLACES, *Introduction: Public order and public security in EU law. Time for Reappraisal*, in *European Papers*, 2024, p. 1318.

⁵³ X. GROUSSOT, G.T. PETURSSON, *Public Security and Public Order in EU Law: Adjudicative Challenges in the Laws of the Internal and Digital Markets*, in *European Papers*, 2024, p. 1387.

⁵⁴ Corte giust. 6 ottobre 2020, cause riunite C-511/18, C-512/18 e C-520/18, *La Quadrature du Net e a.*, punto 135, ECLI:EU:C:2020:791; 5 aprile 2022, C-140/20, *Commissioner of An Garda Síochána*, punto 61, ECLI:EU:C:2022:258; 20 settembre 2022, cause riunite C-793/19 e C-794/19, *SpaceNet*, punto 92, ECLI:EU:C:2022:702.

Tuttavia, deve considerarsi, accanto a quella pubblica, anche una sicurezza privata, che non pare essere esclusa dall'eccezione qui considerata⁵⁵. Del resto, uno degli esempi prospettati dalla Commissione pare confermare ciò: «Using cameras by a supermarket or a bank to detect suspicious customers, for example to conclude that somebody is about to commit a robbery, is not prohibited under Article 5(1)(f) AI Act, when it is ensured that no employees are being tracked and there are sufficient safeguards»⁵⁶. Posto che ci si può legittimamente domandare se questo non contraddica quanto affermato in precedenza nelle linee guida in relazione al fatto che l'eccezione dovrebbe trovare applicazione solo per assicurare «protection of life and health and not to protect other interests, for example property against theft»⁵⁷, rimane la questione di fondo di come intendere la sicurezza in relazione a enti privati e non allo Stato nello svolgimento delle sue funzioni. Diverrà allora essenziale una valutazione alla luce del principio di proporzionalità, come evidenziato dalla stessa Commissione europea.

6. Conclusioni

Le applicazioni dell'IA al riconoscimento delle emozioni possono essere viste come uno sviluppo tecnologico estremamente affascinante e, al tempo stesso, molto controverso. Posto che i rischi per la tutela dei diritti fondamentali – soprattutto, quanto alla protezione dei dati personali – non possono essere sottovalutati, una questione che si ritiene potrà dare luogo a contrasti interpretativi e controversie riguarda la portata delle deroghe previste al divieto di immettere sul mercato, mettere in servizio o usare sistemi di IA per inferire le emozioni di una persona fisica nell'ambito del luogo di lavoro e degli istituti di istruzione. Infatti, i riferimenti ai motivi medici e di sicurezza paiono troppo ampi per essere fatti oggetto di un'interpretazione restrittiva, come dovrebbe effettivamente avvenire per delle eccezioni.

⁵⁵ Condivide questa posizione G.M. DÍAZ GONZÁLEZ, *Prohibited Artificial Intelligence Practices (Chapter II)*, in A. HUERGO LORA (ed.), *The EU Regulation on Artificial Intelligence: A Commentary*, Milano, 2025, p. 60, che afferma: «Permitting the use of such systems by both public and private operators appears unquestionable».

⁵⁶ Annex, cit., p. 85-86.

⁵⁷ Annex, cit., punto 258.



DARIA PERRONE*

SISTEMI INFORMATICI ESPERTI (SIE) E GESTIONE DEL RISCHIO NELL'ACCERTAMENTO DELLA RESPONSABILITÀ PENALE

SOMMARIO: 1. Introduzione: la prevedibilità dell'evento infausto nella c.d. "società del rischio". – 2. Il contesto gnoseologico di partenza, ovvero di come la regolarità di contesto condizioni la prevedibilità dell'agire. – 2.1. Il contesto decisionale ad elevata regolarità epistemica. – 2.2. Il contesto decisionale empiricamente ambiguo. – 2.3. Il contesto decisionale caratterizzato dall'assenza di regole cautelari. – 3. Il ricorso ai *software* predittivi per l'accertamento della colpa nei contesti decisionali irregolari. – 3.1. I *software* predittivi e l'AI Act del 2024: compatibilità e potenziali aspetti critici. – 3.2. Ulteriori criticità: i problemi di attendibilità degli *output* e di fallacità del *software*. 4. Risorse ed incognite: un delicato bilanciamento tra oggettività di giudizio e opacità del processo decisionale.

1. Introduzione: la prevedibilità dell'evento infausto nella c.d. "società del rischio"

La rapida e crescente evoluzione delle conoscenze scientifiche ha portato specularmente all'aumento esponenziale d'inedite categorie di rischio¹, spesso caratterizzate da trame causali «multifattoriali e di lunga latenza, difficilmente assimilabili ai modelli deterministici»². L'"illusione del controllo" propria della società silvo-pastorale e della prima modernità, che si fondava sulla pretesa di conoscere e tipizzare i pericoli derivanti dal vivere comune, ha progressivamente lasciato spazio allo scetticismo nomologico tipico dell'attuale "società del rischio" (o *Risikogesellschaft*, per usare la fortunata espressione di U. Beck)³.

* Ricercatrice di diritto penale, Univ. E-Campus.

¹ V. G. DE FRANCESCO, *Introduzione al convegno*, in G. DE FRANCESCO, G. MORGANTE (a cura di), *Il diritto penale di fronte alle sfide della "società del rischio". Un difficile rapporto tra nuove esigenze di tutela e classici equilibri di sistema*, Torino, 2017, p. IX, il quale avverte che il paradigma del rischio è «consegnato alla temperie evolutiva quanto mai incerta e mutevole della postmodernità».

² D. CASTRONUOVO, *Il caso Eternit. Un nuovo paradigma di responsabilità penale per esposizione a sostanze tossiche?*, in L. FOFFANI, D. CASTRONUOVO (a cura di), *Diritto penale dell'economia, II, Impresa e sicurezza*, Bologna, 2015, p. 107 ss.

³ U. BECK, *Risikogesellschaft. Auf dem Weg in eine andere Moderne*, Frankfurt am Main, 1986, trad. it. W. PRIVITERA, C. SANDRELLI (a cura di), *La società del rischio. Verso una seconda modernità*, Roma, 2000.

Oggi, con la crescente presa d'atto della complessità e della precarietà delle conoscenze disponibili, gli *stakeholders* sono chiamati a fronteggiare rischi sempre più «oscuri»⁴. Così, a fronte un quadro epistemologico eterogeneo, e talvolta persino contraddittorio, affastellato di regole cautelari e di dati empirici, gli agenti sono chiamati a compiere nell'ambito della vita sociale e professionale di ogni giorno difficili valutazioni prognostiche in termini di *risk assesment*, con la conseguente assunzione della responsabilità giuridica per le conseguenze delle proprie scelte.

Ormai, data l'ampiezza del fenomeno, possiamo dire che l'imponderabile è diventato l'orizzonte globale entro cui gli operatori economici e, più in generale i consociati, sono chiamati a confrontarsi. Si dovrebbe parlare, dunque, di un'*imprevedibilità endemica*, in cui sempre più sovente il problema nevralgico - sul piano gestionale e, conseguentemente, giuridico - è quello dell'individuazione delle sfere di responsabilità (o dei capri espiatori, per i più scettici) e, per gli enti, della predisposizione dei *compliance programs*⁵. Non è certo un caso che negli Stati Uniti da anni si discuta del concetto di "*product liability*", in riferimento alla c.d. responsabilità da prodotto (alimentare, farmaceutico, cosmetico, ecc.)⁶, destinata a coinvolgere - a vari livelli - i processi produttivi ed i diversi soggetti partecipanti alla filiera commerciale, attraverso l'individuazione di un'articolata "rete di garanti" (dal produttore, all'importatore, al distributore, e così via).

Ebbene, pur prendendo atto della complessità di un simile scenario, proveremo ai fini della presente indagine a tracciare alcune significative variabili in grado di influire in sede penale nell'accertamento colposo: infatti, come vedremo *infra* § 2, a seconda del diverso contesto decisionale in cui si inserisce la condotta dell'agente, muta sensibilmente il grado di affidabilità del giudizio prognostico e, di conseguenza, la prevedibilità dell'eventuale evento infausto. Procederemo, dunque, analizzando le diversità previsionali sussistenti - da una prospettiva "interna", cioè dal punto di vista dell'agente - a seconda della tipologia di

⁴ C. PIERGALLINI, *Colpa (diritto penale)*, in *Enc. dir.*, X, 2017, p. 222 ss. Sul rapporto tra rischio e pericolo, v. A. GARGANI, *Il rischio nella dinamica dei reati contro l'incolumità pubblica e nei reati di pericolo astratto*, in G. DE FRANCESCO, G. MORGANTE (a cura di), *Il diritto penale di fronte alle sfide della "società del rischio"*, cit., p. 12 ss.; V. MILITELLO, *Diritto penale del rischio e rischi del diritto penale fra scienza e società*, in C.D. SPINELLIS, N. THEODORAKIS, E. BILLIS, G. PAPADIMITRAKOPOULOS (a cura di), *Europe in Crisis: Crime, Criminal Justice, and the Way Forward, Essays in Honour of N. Courakis*, Athens, 2017; C. PERINI, *Il concetto di rischio nel diritto penale moderno*, Milano, 2010, p. 367 ss.

⁵ Il problema è ben noto e diffusamente trattato in tema di *corporate governance*. Si v. N. ABRIANI, *La corporate governance nell'era dell'algoritmo. Prolegomeni a uno studio sull'impatto dell'intelligenza artificiale sulla corporate governance*, in *Nuovo dir. soc.*, 2020, p. 261 ss.; S. A. CERRATO, G. PEIRA, *Risikogesellschaft e corporate governance: prolegomeni sulla costruzione degli assetti organizzativi per la prevenzione dei rischi. Il caso delle imprese agroalimentari*, in *Riv. soc.*, 2019, p. 79 ss.; E. BARCELONA, *La sustainable corporate governance nelle proposte di riforma del diritto europeo: a proposito dei limiti strutturali del c.d. stakeholderism*, in *Riv. soc.*, 2022, p. 1 ss.; M. CATINO, *Miopia organizzativa. Problemi di razionalità e previsione nelle organizzazioni*, Bologna, 2009; S. A. CERRATO, F. CULASSO, E. CROCCO, "Handle with care". *Per una governance dell'intelligenza artificiale nell'impresa: rischi, tecniche di gestione, assetti cibernetici*, in *Riv. soc.*, 2023, p. 448 ss.; A. GARGANI, *Posizioni di garanzia nelle organizzazioni complesse: problemi e prospettive*, in *Riv. trim. dir. pen. econ.*, 2017, p. 508 ss.

⁶ Riflette sul tema C. PIERGALLINI, *Attività produttive e imputazione per colpa: prove tecniche di diritto penale del rischio*, in *Riv. it. dir. proc. pen.*, 1997, p. 1483 ss.; ID., *Danno da prodotto e responsabilità penale. Profili dommatici e politico-criminali*, Milano, 2004 e, successivamente ID., *Colpa e attività produttive: un laboratorio di diritto "cedevole"*, in *Criminalia*, 2014, p. 387 ss. Sotto il profilo etico, tra responsabilità del produttore e rischio consentito, v. A. AMIDEI, *Intelligenza artificiale e responsabilità da prodotto*, in U. RUFFOLO (a cura di), *Intelligenza artificiale*, cit., 124 ss.; B. FRAGASSO, *Intelligenza artificiale e crisi del diritto penale d'evento: profili di responsabilità penale del produttore di sistemi di I.A.*, in F. BASILE, M. BIASI, L. C. CAMALDO, G. CANESCHI, B. FRAGASSO, D. MILANI (a cura di), *Intelligenza Artificiale. Diritto, giustizia, economia ed etica*, Torino, 2025, 37 ss.

contesto gnoseologico disponibile, per poi valutare vantaggi e criticità del ricorso ai Sistemi Informatici Esperti (d'ora in avanti, SIE) per attenuare i *deficit* previsionali propri dei contesti più problematici ed incerti.

2. Il contesto gnoseologico di partenza, ovvero di come la regolarità di contesto condizioni la prevedibilità dell'agire

In generale, dalla prospettiva penalistica, per poter «radicare la colpa all'interno della sfera di libertà dell'agente»⁷, è essenziale che l'aspettativa previsionale sia calata in un preciso contesto gnoseologico spazio-temporale, in base alle regole cautelari disponibili *ex ante* al momento della condotta. L'accertamento della colpa si collega, infatti, al livello di prevedibilità dell'evento infausto al momento dell'agire: è proprio in tale momento che l'agente si trova nella condizione di dover effettuare una scelta comportamentale basandosi sulla *prognosi* di un accadimento futuro e incerto. Breve: l'effettiva rimproverabilità dell'agente dipende dall'inescusabilità *ex ante* del suo difetto previsionale.

Ai fini all'accertamento della colpa, comunque, una differenza significativa dipende dalla natura elettiva o meno delle attività poste in essere negli ambiti che presentano un elevato coefficiente di pericolosità (come avviene nel caso di un'attività sportiva estrema). Nei contesti elettivi, l'agente si espone volontariamente ad un rischio, tanto che, nel caso di condanna, si dovrebbe più propriamente parlare di una c.d. colpa per assunzione.

L'accertamento assume, invece, tratti peculiari allorché l'attività pericolosa non sia elettiva, come espressione dell'esercizio di una libertà, ma si presenti piuttosto come doverosa, ricorrendo sull'agente un obbligo giuridico di intervenire: viene in mente in proposito l'attività posta in essere dagli operatori della Protezione civile, laddove occorre fronteggiare rischi derivanti da fattori esterni ("rischio esogeno"), come appunto alluvioni, terremoti, frane ed altre calamità naturali. In questi casi, gli operatori non solo hanno l'obbligo di essere prudenti e diligenti, ma anche di esserlo "prontamente", non potendo avvalersi della regola dell'astensione⁸. A differenza delle attività elettive⁹, che sono caratterizzate da un rischio endogeno, volontariamente assunto e causato dall'agente stesso; in questi casi, il titolare della posizione di garanzia è stretto tra due doveri, il dovere di

⁷ G. CIVELLO, *Il fondamento epistemico-volitivo della responsabilità colposa: nuovi orizzonti dommatici per la concezione normativa della colpa*, in G. DE FRANCESCO, G. MORGANTE (a cura di), *Il diritto penale di fronte alle sfide della "società del rischio"*, cit., p. 167.

⁸ Di fronte a contesti simili, spesso è lo stesso legislatore che tende a costruire dei modelli di incriminazione a tutela anticipata, secondo lo schema dei reati di pericolo astratto, soprattutto nell'ambito dei delitti contro la salute pubblica, attraverso la previsione di soglie "ipercautelative" di gestione del rischio, in un'ottica apertamente precauzionistica: v. F. D'ALESSANDRO, *Pericolo astratto e limiti-soglia. Le promesse non mantenute del diritto penale*, Milano, 2012, p. 45 ss.; G. DE FRANCESCO, *Colpa e prevenzione del rischio nel campo delle malattie professionali*, in *Dir. pen. proc.*, 2012, p. 665 ss.; G. FORTI, "Accesso" alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione, in *Discrimen*, 19.12.2018; A. GARGANI, *Il rischio nella dinamica dei reati contro l'incolumità pubblica e nei reati di pericolo astratto*, cit., p. 12 ss.

⁹ Con riferimento alle attività pericolose elettive, e segnatamente all'attività sportiva sciistica, v. le riflessioni di D. NOTARO, *Le insidie della colpa nella gestione di attività pericolose lecite. La predisposizione delle pratiche ludico-sportive*, in *Discrimen*, 27.4.2020.

intervenire e quello di non cagionare eventi infausti (c.d. “effetto tenaglia”)¹⁰. È, dunque, comprensibile che – pur percependo la rischiosità della propria condotta – sia maggiormente incline all'azione, incombendo sull'agente un dovere di intervento tempestivo ed efficace: tutti elementi che il giudice, in sede di prognosi postuma, dovrebbe tenere in considerazione.

Tanto premesso, con l'obiettivo di offrire un inquadramento delle diverse tipologie del c.d. *contesto gnoseologico di partenza*, proporremo una distinzione basata, in una progressiva *escalation* di difficoltà previsionale, sulla regolarità epistemica delle norme cautelari disponibili al momento della condotta.

2.1. Il contesto decisionale ad elevata regolarità epistemica

La prima ipotesi – per così dire “idealtipica” – ricorre allorché la condotta sia posta in essere in un contesto decisionale in cui esistono delle regole cautelari ad elevata regolarità epistemica (si pensi, ad esempio, al caso di operazioni mediche di *routine* in assenza di complicanze ed anomalie). Il contesto decisionale in cui si inserisce la condotta è scientificamente affidabile: la sussistenza di regole cautelari comprovate dovrebbe consentire all'agente di raggiungere un certo livello di sicurezza sul fatto che – rispettando tali regole – in sede penale non gli potrà essere mosso alcun rimprovero *ex post* per l'eventuale evento infausto che dovesse verificarsi in conseguenza della sua condotta. L'agente dovrebbe essere in grado di raggiungere quella che nel linguaggio statistico è conosciuta come la c.d. *ability to assign meaningful probabilities*¹¹.

Purtroppo, però, nell'applicazione giurisprudenziale, non può dirsi sempre garantita la corrispondenza “rispetto delle regole cautelari - esclusione della colpa”. Sul piano della responsabilità penale, talvolta, la latitudine del giudizio di colpevolezza trascende la tradizionale logica del *versari in re illicita*, per estendersi anche alle attività realizzate nel pieno rispetto delle *guidelines* e delle *best practices* specificamente adottate nei settori di riferimento. Così, ad es., nel settore ambientale, di fronte a fenomeni di contaminazione a seguito dello stratificarsi nel tempo di attività di tipo industriale, la giurisprudenza ha ritenuto sussistente la colpa, nonostante il rispetto del rischio consentito in base alle leggi scientifiche disponibili al momento della condotta. Emblematico, in proposito, è il caso della centrale termoelettrica “Tirreno Power” di Vado Ligure¹²: nonostante la conformità delle emissioni ai valori-soglia, il giudice ha riconosciuto, sia pure in sede cautelare, la configurabilità degli illeciti di cui agli artt. 434, 589 e 590 c.p., in quanto i valori-soglia sono stati ritenuti *ex post* non adeguati ai più rigorosi *standard* cautelativi successivamente introdotti a livello europeo.

2.2. Il contesto decisionale empiricamente ambiguo

La seconda ipotesi ricorre allorché l'agente si trovi a dover effettuare una scelta comportamentale in un contesto gnoseologico ambiguo e contraddittorio, in cui – in assenza

¹⁰ Per una lucida analisi di tale effetto, v. F. GIUNTA, *Quale colpa per la Protezione civile?*, in *La giustizia penale*, 2016, II, p. 126 ss., il quale riflette sulla possibilità di limitare in tali casi l'addebito colposo ai soli casi di colpa grave, similmente a quanto avviene in tema di responsabilità medica, a seguito del c.d. decreto Balduzzi.

¹¹ Per approfondimenti, v. I. GLETTE-IVERSEN, R. FLAGE, *On unpredictable events in risk analysis*, in *Safety Science*, 2024.

¹² Il caso è riportato da A. GARGANI, *Jus in latenti. Profili di incertezza del diritto penale dell'ambiente*, in *Crim.*, 2019, p. 111 ss. Sul tema del rischio consentito rispetto al diritto penale ambientale, senza alcuna pretesa di esaustività, v. C. RUGA RIVA, *Il caso Ilva: disastro e avvelenamento dolosi*, in L. FOFFANI – D. CASTRONUOVO (a cura di), *Casi di diritto penale dell'economia*, II, Bologna, 2015, p. 160 ss.; S. ZIRULIA, *Esposizione a sostanze tossiche e diritto penale*, Milano, 2018, p. 410 ss.

di regole cautelari consolidate o comunque che non superano i criteri di attendibilità tipici della falsificazione popperiana – si registrino dati discordanti.

In simili contesti decisionali, l'agente difficilmente potrebbe essere in grado di formulare una prognosi attendibile circa la propria rimproverabilità soggettiva rispetto ai possibili (e futuri) effetti dannosi o pericolosi collegati al suo agire¹³: si parla, in questo senso, di una *lack of ability to establish accurate models or cause-effect relationships*¹⁴.

Si pensi in proposito agli eventi disastrosi che si verificano a seguito di calamità naturali. Spesso, a causa della difficoltà di effettuare una prognosi attendibile rispetto al verificarsi di eventi disastrosi quali frane, alluvioni, inondazioni, la scienza non è in grado di formulare regole cautelari generalmente condivise e universalmente accreditate. In relazione, ad es., all'alluvione di Sarno del 1998, il sindaco del Comune fu condannato per il reato di omicidio colposo (plurimo) per aver omesso di dare il tempestivo allarme e di disporre l'evacuazione delle zone a rischio e, in concorso con un assessore, per aver fornito notizie imprudentemente rassicuranti alla popolazione in pericolo. Annullando con rinvio la sentenza d'appello che aveva assolto il sindaco per imprevedibilità dell'evento, la Cassazione¹⁵ ritenne «in mancanza di leggi scientifiche che consentano di conoscere preventivamente lo sviluppo di eventi naturali calamitosi» che il giudizio di prevedibilità vada compiuto tenendo conto di quello che potrebbe accadere in futuro «nel senso che involge un giudizio di rappresentabilità di possibili, ulteriori e più gravi eventi dannosi (...) senza ignorare l'esistenza di una possibilità di evoluzione del fenomeno e ipotizzando quindi la più distruttiva ipotesi che potesse verificarsi o che il fenomeno disastroso poteva comportare»¹⁶: il riferimento – invero “confuso”¹⁷ – è qui alla prevedibilità del c.d. *worst case scenario*.

In queste ipotesi, un dato che la giurisprudenza sembra ritenere decisivo per fondare il giudizio colposo, come ad es. nel caso ThyssenKrupp¹⁸, è rappresentato dal verificarsi dei c.d. eventi sentinella o “*near misses*”. Gli eventi premonitori, intesi come *precursor signals* o *warnings*, dovrebbero indurre nell'agente-modello quanto meno il sospetto della pericolosità

¹³ Per una disamina delle problematiche in simili contesti incerti, v. A. GARGANI, *La “flessibilizzazione” giurisprudenziale delle categorie classiche del reato di fronte alle esigenze di controllo penale delle nuove fenomenologie di rischio*, in *Leg. pen.*, 2011, p. 422 ss.; A. ORSINA, *Rischio da incertezza scientifica e modelli di tutela penale*, Torino, 2015; A. PERIN, *La crisi del “modello nomologico” fra spiegazione e prevedibilità dell'evento nel diritto penale*, in *Riv. it. dir. proc. pen.*, 2014, p. 1371 ss.; ID., *Prudenza, dovere di conoscenza e colpa penale. Proposta per un metodo di giudizio*, Napoli, 2020.

¹⁴ V. I. GLETTE-IVERSEN, R. FLAGE, *On unpredictable events in risk analysis*, cit., alla cui bibliografia si rimanda.

¹⁵ Cass., sez. IV, 12.3.2010, n. 16761, in *Foro it.*, 2011, p. 482 ss.: la vicenda trae origine dall'alluvione che colpì il Comune di Sarno in Campania. Il Sindaco fu condannato (dopo essere stato assolto nei giudizi di merito) dalla Cassazione alla pena di cinque anni di reclusione per il delitto di omicidio colposo plurimo.

¹⁶ Cass., sez. IV, 12.3.2010, n. 16761, cit.

¹⁷ Così A. PERIN, *Prudenza, dovere di conoscenza e colpa penale*, cit., p. 178 ss. Sempre in senso critico, v. anche le riflessioni di F. GIUNTA, *Relazione*, in AA.VV., *La protezione civile nella società del rischio*, Bologna, 2019, 69, p. 71 ss., secondo cui la Cassazione, con la sentenza *de quo*, avrebbe elaborato la tesi della “prevedibilità in astratto”, sostenendo che si sarebbe dovuto prevedere l'alluvione in tutte le zone che in astratto potevano essere colpite. La tesi della prevedibilità in astratto, nel diverso settore della responsabilità per le malattie professionali, è già stata debitamente censurata da C. PIERGALLINI, *Colpa e attività produttive*, cit., p. 383, secondo cui «il disvalore di condotta non viene più concepito come una rappresentazione causale anticipata, di taglio nomologico, ma viene soppiantata da una sfumata non escludibilità dell'evento, o dalla rappresentabilità di un generico evento di danno alla salute, addirittura ricavabile da diafani elementi fattuali».

¹⁸ V. sul punto Cass. S.U. 24.4.2014, n. 38343, ThyssenKrupp, in *Cass. pen.*, 2015, 2, p. 426, relativa al noto caso del disastroso incendio divampato nel dicembre 2007 nell'acciaieria torinese di proprietà della multinazionale tedesca ThyssenKrupp, a seguito del quale morirono sette operai. Con sentenza di primo grado (sentenza poi riformata in appello ed in Cassazione), l'amministratore delegato era stato condannato per omicidio doloso, mentre agli altri garanti era mosso esclusivamente un rimprovero di colpa (art. 589 c.p.), sia pure aggravata.

della propria condotta, pur ancora non supportata da regole cautelari consolidate. In questi casi, il livello previsionale del c.d. agente modello (il tanto vituperato *homo eiusdem conditionis et professionis*) andrebbe commisurato su di un bagaglio conoscitivo dal contenuto prettamente empirico, sfruttando gli episodi pregressi per trarre dalla situazione concreta una proiezione di pericolosità della condotta. Così, ad es., in relazione al disastro verificatosi nel Vajont nel 1963, in cui già da tempo si erano registrati diversi episodi “sentinella”, la Cassazione¹⁹ ha ritenuto integrato il reato di disastro ambientale aggravato dalla previsione dell'evento.

2.3. Il contesto decisionale caratterizzato dall'assenza di regole cautelari

Un ulteriore contesto decisionale si profila di fronte al rischio imprevedibile, inteso come di *irreducible uncertainty*, allorquando non sussistano regole cautelari di riferimento. Si pensi alla realizzazione di attività del tutto innovative, in cui gli studi e le ricerche non hanno ancora superato lo stadio teorico e non esistono protocolli di riferimento, come ad es. in relazione alla pratica degli *xenotrapianti* o alla commercializzazione di prodotti transgenici.

Ed ancora, senza necessariamente fare riferimento alle avveniristiche frontiere della scienza sperimentale, si pensi ai danni prodotti a seguito dei terremoti²⁰: ad oggi, purtroppo, per quanto la scienza si avvalga di strumenti d'avanguardia sempre più raffinati, non è possibile elaborare delle previsioni attendibili dell'attività sismica (sul quando, sul dove e sull'intensità del fenomeno). Vengono in mente, ad es., il terremoto di Amatrice del 2016²¹ o quello dell'Aquila del 2009, in cui la Corte d'Appello²² ribaltò il giudizio di primo grado, ritenendo che nessuna censura potesse essere mossa, secondo un giudizio *ex ante*, agli imputati coinvolti.

In simili contesti d'irriducibile incertezza, laddove cioè non sussistano regole cautelari cui parametrare il comportamento doveroso dell'agente modello, l'agente si troverebbe ad effettuare una scelta comportamentale in una situazione di mera latenza del rischio²³, trattandosi di una situazione in cui il verificarsi dell'evento infausto resterebbe nella sua mente solamente ad uno stato sub-previsionale, nel senso che, al momento della condotta, egli potrebbe solo non escluderlo. Si potrebbe parlare al più di una colpa per la violazione di regole ultra-prudenziali riconducibili ad un generico dovere di attenersi al brocardo latino “*in dubiis, abstinere*”.

¹⁹ V. Cass., Sez. IV, 25.3.1971, in *Foro it.*, 94, 12, 1971, p. 717 ss.

²⁰ Il tema dell'imprevedibilità dei disastri sismici è noto: v. A. AMATO, A. CERASE, F. GALADINI (a cura di), *Terremoti, comunicazione, diritto. Riflessioni sul processo alla «Commissione Grandi Rischi»*, Milano, 2015; S.E. HOUGH, *Prevedere l'imprevedibile. La tumultuosa scienza della previsione dei terremoti*, Milano, 2013.

²¹ In relazione al terremoto di Amatrice, nel giugno 2011, con sentenza di primo grado del Trib. Rieti (notizia riportata dal *Corriere della Sera*, 4.6.2011) furono assolti dal reato di disastro colposo ed omicidio colposo, con formula “perchè il fatto non sussiste”, cinque dei sette imputati; mentre per gli altri due, tra i quali il sindaco di Accumoli, l'assoluzione è stata pronunciata “perchè il fatto non costituisce reato”.

²² Corte d'Appello L'Aquila, 10.11.2014, in www.penalecontemporaneo.it, 16.2.2015, con *ivi* nota di A. GALLUCCIO, *La sentenza di appello sul caso del terremoto dell'Aquila*. In commento, v. anche D. NOTARO, *Nesso di causalità e colpa alla prova delle dinamiche “incerte”: osservazioni a margine del processo sul terremoto dell'Aquila*, in *Leg. pen.*, 29.1.2016.

²³ L'espressione è di F. URBAN, *La curatela precauzionale della colpa nella gestione delle calamità naturali: prevedibilità ed evitabilità dell'evento*, in G. DE FRANCESCO, G. MORGANTE (a cura di), *Il diritto penale di fronte alle sfide della “società del rischio”*, cit., 179.

3. Il ricorso ai software predittivi per l'accertamento della colpa nei contesti decisionali irregolari

Come abbiamo visto, in contesti decisionali caratterizzati da un elevato tasso di irregolarità (si pensi ad es. alla commercializzazione di prodotti nocivi per la salute o di nuovi farmaci, oppure alla creazione di nuovi *software* per la protezione o la criptazione dei dati sensibili, all'inserimento nel mercato di nuovi protocolli di comunicazione *wireless*, alle sperimentazioni nel settore aerospaziale, alle implicazioni della messa in commercio di autoveicoli a guida autonoma o assistita, ecc.) è assai difficile prevedere le conseguenze dell'agire, per l'ambiguità o addirittura l'assenza delle regole cautelari di riferimento.

In simili *decision-making context*, caratterizzati da irregolarità epistemica, il ricorso ai SIE potrebbe offrire un valido supporto per l'accertamento della responsabilità colposa, ai fini della valutazione della prevedibilità ed evitabilità dell'evento infausto dalla prospettiva dell'agente in base al suo effettivo bagaglio gnoseologico di partenza²⁴. In questo senso, potrebbe essere utile sondare l'utilizzabilità dei c.d. *software* predittivi, ossia di quei programmi – basati oggi anche su sistemi di AI – in grado di calcolare le probabilità di un determinato evento a partire dall'immissione di una certa quantità di *input* (c.d. *data-set* iniziale)²⁵. In tali SIE, il *data-set* iniziale dovrebbe essere impostato sfruttando le conoscenze disponibili al momento della condotta, in quello che abbiamo definito come “contesto decisionale di partenza”. L'*output* dovrebbe restituire, invece, l'esito previsionale dell'agente modello, ossia il livello di capacità previsionale e di esigibilità cui parametrare la condotta effettivamente posta in essere dall'agente concreto, per valutarne l'eventuale responsabilità colposa.

3.1. I software predittivi e l'AI Act del 2024: compatibilità e potenziali aspetti critici

Per valutare l'utilizzabilità dei *software* predittivi in tale ambito, occorre oggi fare riferimento preliminare al regolamento UE 2024/1689 (cd. AI Act)²⁶, con il quale il Parlamento europeo e il Consiglio hanno disciplinato i criteri per un ricorso ai sistemi di AI sicuro e rispettoso dei diritti fondamentali dell'Unione, distinguendo il livello di cautela a seconda dei diversi gradi di rischio (basso, alto e inaccettabile), in base al c.d. *risk-based approach*²⁷. Così, il regolamento ha delineato una serie di obblighi e requisiti ad intensità

²⁴ Sul tema, senza pretese di completezza, si rimanda a A. BALDASSARRE, *Diritto ex machina? L'intelligenza artificiale e l'attività giurisdizionale*, in R. GIORDANO, A. PANZAROLA, A. POLICE, S. PREZIOSI, M. PROTO (a cura di), *Il diritto nell'era digitale. Persona, Mercato, Amministrazione, Giustizia*, Milano, 2022, p. 461 ss.; C. BURCHARD, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, in *Riv. it. dir. proc. pen.*, 2019, 62, 4, p. 1909 ss.190; G. COMANDÉ, *Intelligenza artificiale e responsabilità tra liability e accountability. Il carattere trasformativo dell'IA e il problema della responsabilità*, in *An. Giur. Econ.*, 2019, p. 169 ss.; S. CREESE, *The Threat from AI*, in D.J. BAKER, P.H. ROBINSON (a cura di), *Artificial Intelligence and the Law. Cybercrime and Criminal Liability*, London-New York, 2021, p. 201 ss.; M. DSOUZA, *Don't Panic: Artificial Intelligence and Criminal Liability*, in D.J. BAKER, P.H. ROBINSON (a cura di), *cit.*; L. LUPÀRIA – G. FIORELLI, *Diritto probatorio e giudizi criminali ai tempi dell'Intelligenza Artificiale*, in R. GIORDANO, A. PANZAROLA, A. POLICE, S. PREZIOSI, M. PROTO (a cura di), *Il diritto nell'era digitale, cit.*, 779 ss.; L. PICOTTI, *Diritto penale, tecnologie informatiche e intelligenza artificiale: una visione d'insieme*, in A. CADOPPI, S. CANESTRARI, A. MANNA, M. PAPA (a cura di), *Cybercrime*, Torino, 2023, p. 32 ss.

²⁵ Sia consentito sul punto fare riferimento a D. PERRONE, *La prognosi postuma tra distorsioni cognitive e software predittivi. limiti e possibilità del ricorso alla "giustizia digitale integrata" in sede di accertamento della colpa*, Torino, 2021.

²⁶ Reg. UE 2024/1689 del 13.6.2024.

²⁷ Così A. MANGIONE, *Intelligenza artificiale, attività d'impresa e diritto penale. La «funzione di garanzia» nell'organizzazione e dell'organizzazione per la «sorveglianza dell'ai»*, Torino, 2024, 195 ss.

crescente, in base ad un modello che è stato correttamente definito come “*regulatory sandbox*”²⁸.

Ebbene, i *software* predittivi per l'accertamento colposo sembrano collocarsi nell'ambito della categoria definita dal regolamento “a rischio elevato”: ai sensi dell'art. 8, lett. a), all. III, sono classificati come “ad alto rischio”, infatti, «i sistemi di IA destinati a essere usati da un'autorità giudiziaria o per suo conto per assistere un'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti, o a essere utilizzati in modo analogo nella risoluzione alternativa delle controversie». Inoltre, ai sensi dell'art. 6, par. 3, Reg. sono sistemi “ad alto rischio”, oltre quelli elencati nell'allegato III, anche quelli che presentano un rischio significativo di danno per la sicurezza o i diritti fondamentali delle persone fisiche, anche nel senso di non influenzare materialmente il risultato del processo decisionale.

Se, dunque, tali *software* predittivi rientrano nella categoria ad alto rischio, il regolamento europeo impone l'osservanza di alcuni fondamentali requisiti (art. 8), oltre all'applicazione di preventive valutazioni di conformità, d'impatto del rischio, di registrazione e monitoraggio del funzionamento (artt. 9-12). Per essi è previsto, inoltre, l'applicazione di un sistema obbligatorio di tracciamento, trasparenza e soprattutto sorveglianza umana, ex artt. 13 e 14 («*human in the loop*»).

Proprio nell'applicazione di quest'ultimo requisito della sorveglianza umana, che sembra evocare l'irrealistica pretesa di attribuire ad uno o più soggetti una posizione di garanzia con obblighi di sorveglianza misti ad obblighi di controllo sull'operato dell'IA, è possibile intravedere delle potenziali criticità. Pare in effetti difficile ipotizzare che una o più persone fisiche abbiano la reale capacità di “sorvegliare” l'operato di una *machine learning*, a meno che tale sorveglianza non venga interpretata come non molto di più di un'operazione meramente formale, di verifica “esterna” sul funzionamento dell'algoritmo. Il sospetto è che per tal via si finisca per “scaricare” sull'uomo gli eventuali errori – valutati *ex post* – della macchina, finendo per condurre «alla edificazione di una responsabilità omissiva per il mancato impedimento dell'evento avverso prodotto dall'agente artificiale»²⁹.

3.2. Ulteriori criticità: i problemi di attendibilità degli output e di fallacità del software

L'utilizzabilità concreta di SIE in contesti decisionali caratterizzati da un'elevata irregolarità epistemica pone, purtroppo, anche ulteriori problemi, che – per ragioni di sintesi – possiamo in questa sede solo accennare. Oltre alle già accennate criticità legate – come richiesto all'AI Act – alla necessità di una sorveglianza “umana”, ulteriori difficoltà sono legate alla trasparenza del meccanismo decisionale ed all'attendibilità degli *output* e di fallacità del *software*³⁰.

²⁸ Ci si riferisce a A. MERLINO, *Regulatory sandbox: una nuova prospettiva ordinamentale*, Napoli, 2022.

²⁹ F. CONSULICH, *Flash offenders. Le prospettive di accountability penale nel contrasto alle intelligenze artificiali devianti*, in *Riv. it. dir. proc. pen.*, 1049. In senso simile, v. anche A. CAPPELLINI, *Reati colposi e tecnologie dell'intelligenza artificiale*, *Reati colposi e tecnologie dell'intelligenza artificiale*, in G. BALBI, F. DE SIMONE, A. ESPOSITO, S. MANACORDA (a cura di), *Diritto penale e intelligenza artificiale. “Nuovi scenari”*, Torino, 2022, p. 30 ss.; A. GIANNINI, *Intelligenza artificiale, human oversight e responsabilità penale: prove d'impatto a livello europeo*, in *Discrimen*, 21.11.2021, p. 16 ss.; A. MANGIONE, *Intelligenza artificiale*, cit., p. 195 ss.; L. PICOTTI, *Categorie tradizionali del diritto penale e intelligenza artificiale: crisi o palinogenesi?*, in *Sist. pen.*, 31.7.2024.

³⁰ Gli aspetti critici di tali sistemi sono variamente affrontati, anche sotto il profilo etico e di tutela dei diritti umani, da numerosi Autori, tra cui: Y. BATHAEE, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, in *Harvard Journ. Law & Tech.*, 2018, 31, 2, p. p. 889 ss.; J. BURELL, *How the Machine Thinks: Understanding Opacity in Machine Learning Algorithms*, in *Big Data and Society*, 15.9.2015; A. D. CARRERAS, *The Moral (Un)intelligence*

Sotto il primo aspetto, la trasparenza del meccanismo decisionale, che dovrebbe costituire una garanzia fondamentale del giusto processo per poter verificare in contraddittorio tra le parti la correttezza degli *output*, difficilmente potrebbe essere assicurata. Infatti, il funzionamento del *software* – se prodotto da un privato – potrebbe essere coperto da *copyright*: per ovviare a ciò in dottrina³¹ si è ipotizzata la tesi (in verità difficilmente praticabile e, infatti, non accolta nella giurisprudenza statunitense) di vietare l'opposizione del segreto industriale sui dati che integrano il c.d. codice sorgente allorché lo strumento predittivo sia impiegato da pubbliche autorità, come quelle giudiziarie in un processo penale.

Per quanto riguarda il secondo aspetto, invece, il problema nasce dal fatto che l'attendibilità degli *output* dipende a monte dalla selezione degli *input*: più elevato è il numero di dati disponibili, maggiore è la loro affidabilità³². Viceversa, minore sono i dati disponibili, minore è l'affidabilità, secondo il principio statistico *garbage in – garbage out* (GIGO). Ciò vuol dire che è possibile progettare un *software* predittivo realmente affidabile solo laddove gli sviluppatori dispongano di una vasta mole di dati (come ad es. nei contesti regolari) in ordine al modo di funzionamento delle norme cautelari. Nei contesti decisionali caratterizzati da irregolarità epistemica, invece, come abbiamo visto, le regole cautelari sono ambigue, contraddittorie e talvolta persino assenti: in definitiva, proprio laddove ci sarebbe più bisogno dei *software* predittivi, è più probabile che questi si rivelino inattendibili, fallaci e, in ultima istanza, inutili.

4. Risorse ed incognite: un delicato bilanciamento tra oggettività di giudizio e opacità del processo decisionale

La società civile sembra riporre aspettative sempre più elevate negli strumenti digitali di nuova generazione. In particolare, l'ideale di una “giustizia digitale”, basata sul puro calcolo matematico e depurata da ogni emotività umana, presenta da sempre un'indubbia suggestività, realizzando quella primordiale aspirazione ad una giustizia “esatta” già teorizzata dal “*calculus*” di Leibniz. I sistemi di calcolo automatizzati sembrano prometterci

Problem of Artificial Intelligence in Criminal Justice: A Comparative Analysis under Different Theories of Punishment, in *Ucla J. Law & Tech.*, 30.12.2020; N. CRISTIANINI, *Machina sapiens. L'algoritmo che ci ha rubato il segreto della conoscenza*, Bologna, 2024; L. D'AVACK, *La rivoluzione tecnologica e la nuova era digitale: problemi etici*, in U. RUFFOLO (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milano, 2020; p. 3 ss.; E. FAZIO, *Intelligenza artificiale e diritti della persona*, Milano, 2023; G. FINOCCHIARO, *The Regulation of Artificial Intelligence*, in *AI & Society*, 3.4.2023; A. FIORELLA, *Responsabilità penale del Tutor e dominabilità dell'Intelligenza Artificiale. Rischio permesso e limiti di autonomia dell'Intelligenza Artificiale*, in R. GIORDANO, A. PANZAROLA, A. POLICE, S. PREZIOSI, M. PROTO (a cura di), *Il diritto nell'era artificiale*, cit., p. 651 ss.; L. FLORIDI, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Milano, 2022; M. E. FLORIO, *Il dibattito sulla responsabilità penale diretta delle LA: “molto rumore per nulla”?*, in *Sist. pen.*, 8.2.2024; R. KOSTORIS, *Intelligenza artificiale, strumenti predittivi e processo penale*, in *Cass. pen.*, 2024, 5, p. 1642 ss.; V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in U. RUFFOLO (a cura di), *Intelligenza artificiale*, cit.; A. M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale: una sfida tra evidence based practices e tutela dei diritti fondamentali*, in *Arch. pen.*, 2021, 1, p. 1 ss.; A. POLICE, *Scelta discrezionale e decisione algoritmica*, in R. GIORDANO, A. PANZAROLA, A. POLICE, S. PREZIOSI, M. PROTO (a cura di), *Il diritto nell'era digitale*, cit., p. 493 ss.

³¹ Così, v. S. QUATTROCOLO, *Artificial intelligence, Computational Modelling and Criminal Proceedings. A Framework for a European Legal Discussion*, Springer, 2020, p. 95 ss., la quale ritiene che una simile soluzione rappresenti una condizione necessaria per poter esperire un controllo indipendente sui dati forniti dal sistema.

³² In questo senso, v. P. GARBOLINO, *Nuovi strumenti logici ed informatici per il ragionamento giudiziario: le Reti Bayesiane*, in *Cass. pen.*, 2007, 47, 326 ss., alla cui bibliografia sul punto si rimanda.

quell'ideale di giustizia che promana "dall'alto", che, in un certo senso, sovrasta la fallibilità dell'uomo, condizionato da pregiudizi e da altre imponderabili variabili soggettive³³. Il "mondo degli algoritmi" sembra così suggerire la possibilità di «sterilizzare l'assetto entropico»³⁴ della post-modernità, sostituendo al disordine che la caratterizza la promessa di una giustizia certa ed infallibile.

Eppure, l'utilizzabilità dei *software* predittivi per l'accertamento colposo in ambito penale si espone – almeno per il momento – a numerose criticità, in termini soprattutto di attendibilità degli *output* e di trasparenza decisionale.

Ciò non vuol dire che le potenzialità dischiuse dalle nuove tecnologie non debbano continuare ad essere esplorate. I progressi scientifici e tecnologici compiuti nell'elaborazione di tali *software* disvelano per il futuro enormi potenzialità ed inediti ambiti di applicazione: non è da escludere che prossimamente tali *tools* possano giocare un ruolo decisivo nel processo penale specialmente in quei contesti di rischio fortemente aleatori – caratterizzati da elevati margini di incertezza, come quello dei trattamenti farmacologici sperimentali o della lavorazione delle sostanze pericolose – in cui la fase della mappatura del rischio risulta più complessa e difficoltosa per l'uomo. Per queste ragioni, riteniamo che una posizione aperta alle possibili forme di convivenza e collaborazione fra decisore umano e SIE possa condurre a risultati virtuosi e fecondi: in futuro, nel rispetto della garanzia di trasparenza del meccanismo decisionale e di attendibilità degli *output*, la progettazione di *software* predittivi, volti a sopperire ai limiti umani, potrebbe fungere da strumento di supporto alla valutazione giudiziale, pur senza sostituirsi ad essa.

³³ Così, S. QUATTROCOLO, *Qualcosa di meglio del diritto (e del processo) penale*, in *Discrimen*, 26.6.2020, la quale si interroga se l'analisi computazionale dei precedenti attraverso i *software* predittivi sia "qualcosa di meglio" rispetto alla discrezionalità del giudice.

³⁴ R. KOSTORIS, *Intelligenza artificiale, strumenti predittivi e processo penale*, cit., p. 1642 ss.



ANDREA CAPUTO*

GIUDICE-MACCHINA E DIGNITÀ UMANA: *ERROR 404*

SOMMARIO: 1. Introduzione. – 2. Ragionamento umano e calcolo algoritmico. – 3. La dimensione relazionale della dignità. – 4. Giudice-macchina e dignità umana: *Error 404*

«Un contratto, un delitto, un processo sono degli uomini uno di fronte all'altro.
Vuol dire che bisogna capire quegli uomini per capire il diritto.
Ma questa è materia ribelle ai numeri e anche alle parole».

F. CARNELUTTI, *Matematica e diritto*, in *Rivista di diritto processuale*, 1, 1951, p. 211 s.

1. Introduzione

La progressiva integrazione di sistemi dotati della cosiddetta Intelligenza Artificiale nelle pratiche umane è un fenomeno ormai consolidato, che procede con velocità e intensità crescenti man mano che il progresso tecnologico avanza con i suoi sbalorditivi risultati.

Di fronte a questi sviluppi v'è chi si entusiasma per i vantaggi che ne seguiranno in termini di efficienza e precisione, e in generale per l'aprirsi di possibilità finora inimmaginabili. Ma v'è anche chi si preoccupa per le conseguenze che questi potranno determinare a diversi livelli, dalle ricadute sul piano occupazionale alla compressione di alcuni principi fondamentali delle società liberal-democratiche fino alle trasformazioni di senso che potrebbero subire le pratiche umane. A rendere il diritto un campo particolarmente sensibile alle preoccupazioni è il fatto ovvio che per esso passa l'esercizio del potere in società, e una sua colonizzazione da parte dell'IA rappresenta lo scenario tipico delle distopie fantascientifiche.

Senza giungere a tanto, è ovvio porsi il problema dei modi in cui l'ingresso dell'IA nel diritto potrebbe trasformare la società. Bisogna, però, condursi su queste acque senza farsi inghiottire da Scilla e Cariddi del tecno-entusiasmo, da un lato, e della tecno-fobia, dall'altro. La teoria del diritto non deve dialogare con speranze o timori, ma con ciò che sappiamo del funzionamento dell'IA.

* Dottore di Ricerca in Filosofia del Diritto; Ricercatore, Fondazione Luigi Einaudi di Torino.

Nel piccolo, questo è ciò che si tenterà di fare in queste pagine prendendo in esame un tema specifico: il modo in cui l'introduzione nei sistemi giurisdizionali di un giudice-macchina (GM) impedirebbe il rispetto della dignità umana. L'argomento procederà come segue. Innanzitutto si metterà a fuoco la distanza tra l'IA e l'intelligenza umana; successivamente si fornirà una presentazione della questione relazionale nell'ambito delle teorie della dignità umana e del modo in cui essa viene in rilievo nella dimensione processuale; infine si vedrà in che modo l'introduzione di un GM sarebbe lesiva della dignità umana.

2. Ragionamento umano e calcolo algoritmico¹

È un errore comune quello di prendere troppo alla lettera la definizione di Intelligenza Artificiale. Certo la semplificazione agevola la comunicazione, ma nel discorso accademico e specialistico (per non parlare di quello politico) è importante tenere a mente che quando si parla di IA si fa uso di un nome impreciso. Riuscendo, soprattutto i Large Language Model (LLM), a farlo ormai magnificamente, i moderni sistemi di IA tendono a far dimenticare che dell'intelligenza umana essi riproducono soltanto l'apparenza. Quando nel 1997 l'algoritmo DeepBlue sconfisse a scacchi il campione del mondo Garry Kasparov, qualcuno celebrò l'avvento di un'IA del tutto equivalente all'intelligenza umana². Se il dibattito non si chiuse allora è perché questa considerazione si fondava su una concezione dell'intelligenza umana gravemente riduttiva. Il confronto era condotto, infatti, soltanto sul piano di quella specifica dimensione dell'intelligenza che si esprime nel calcolo. Come ha scritto Gerd Gigerenzer, la scacchiera è governata dal cosiddetto *principio del mondo stabile*³: le pedine sono limitate a certe mosse e lo scopo prestabilito vincola l'imprevedibilità dell'avversario. Ciò crea un mondo in cui la macchina eccelle grazie alla sua straordinaria capacità di calcolo. Tuttavia, vincendo la partita contro Kasparov, DeepBlue non stava *giocando* a scacchi in senso proprio. Pur partecipando a una pratica umana, DeepBlue non poteva comprenderne il senso, non poteva rappresentarsi il significato degli scacchi come gioco né del gioco come attività sociale particolare. In effetti, a rigore, anche parlare di *partecipazione* della macchina significa utilizzare termini imprecisi. Programmata per applicare le regole a uno scopo definito, la macchina, eseguendo il compito mediante calcoli probabilistici, non aveva *comprensione* del gioco. Del *gioco* e della *partecipazione* a una pratica umana, evidentemente, ciò ha soltanto l'aspetto esteriore.

La moderna IA utilizza il linguaggio allo stesso modo. Essa non ha comprensione dei segni che utilizza ma li adopera eseguendo calcoli probabilistici circa le loro associazioni. Per quanto oggi riesca a farlo sprigionando una potenza di calcolo talmente elevata da consentirle di raggiungere nelle sue operazioni un livello di sofisticazione a dir poco impressionante, l'IA moderna, all'osso, funziona ancora come diceva John Searle nel 1990, processando le informazioni in modo da *trattarle come, contarle come o farle funzionare come 0 e 1*⁴. Nonostante la definizione non renda giustizia del complesso funzionamento specialmente dei sistemi di

¹ L'autore vuole ringraziare l'anonimo *referee* per la sollecitazione a chiarire meglio questo importante punto di partenza argomentativo.

² Cfr. G. GIGERENZER, *Perché l'intelligenza umana batte ancora gli algoritmi*, Milano, p. 54 ss.

³ Cfr. *ivi*, pp. 53 s.

⁴ Cfr. J. R. SEARLE, *Is the Brain a Digital Computer?*, in *Proceedings and Addresses of the American Philosophical Association*, 3, 1990, p. 25.

Deep learning, dietro gli strabilianti risultati prodotti dalle macchine che chiamiamo “intelligenti”, al fondo, c’è questo: per leggere la realtà l’IA non attinge al senso delle pratiche umane, bensì codifica le informazioni in segni matematici: omologhi, fungibili e calcolabili.

Per questo motivo, quando si dice che una macchina *ragiona* si sta facendo, proprio come con il gioco, soltanto un’analogia con le attività umane. La stessa definizione di IA, insomma, è fondata su una mera analogia con l’intelligenza umana. Del resto, come è stato suggerito da uno Luciano Floridi, tra i più autorevoli studiosi in materia, tale imprecisa analogia sarebbe da superare per adottarne una che indichi questi sistemi con un nuovo tipo di *Agency*, piuttosto che di *Intelligence*⁵. È noto l’argomento della “stanza cinese” con cui Searle si opponeva alla tesi sviluppata da Alan Turing con il suo celebre *Imitation Game*⁶. La macchina accede alla realtà umana mediante un’associazione probabilistica di elementi sintattici. In questo, l’IA replica con una straordinaria potenza di calcolo quelle attività mentali umane suscettibili di matematizzazione. Essa, tuttavia, non ha accesso alla dimensione semantica del linguaggio. Il calcolo simula l’atto umano del comprendere, ma in realtà non tiene conto di tutto ciò che dell’esperienza umana eccede la sintassi.

Invero, specialmente attorno ai LLM, studiosi ed esperti tendono oggi a dividersi sul funzionamento, e quindi sulla “natura” della moderna IA⁷. Non tutti concordano sul fatto che i LLM sarebbero, come la stanza cinese di Searle, meri *pappagalli stocastici*. È anzi ben possibile, come è stato notato, che una simile definizione sia ormai da considerarsi quantomeno riduttiva⁸. Nondimeno, gli argomenti che con più insistenza propongono un’associazione forte tra IA e intelligenza umana sono, tendenzialmente, di due tipi. Da un lato si trovano gli argomenti di chi guarda al futuro per sostenere che, quand’anche le macchine oggi esistenti fossero ancora effettivamente limitate alla dimensione sintattica, questo stato di cose è destinato ad essere decisamente superato da più o meno imminenti innovazioni tecnologiche. Argomenti del genere guardano in diverse direzioni: computer quantistici, robot dotati di sistemi percettivi della realtà, e così via. Tuttavia, come si diceva nell’introduzione, sembra sconsigliabile ragionare, specialmente nel campo della riflessione filosofica, su scenari futuri ancora non definiti o del tutto ipotetici. Anche trascurando che, allo stato dell’arte, sembrano esservi limiti fisici oggettivi che impedirebbero il salto delle macchine alla dimensione semantica, il punto è che, in termini di pensiero filosofico-critico, tematizzare tali scenari futuri vale pressappoco quanto prendere in esame gli scenari distopici che vengono qua e là proposti nel dibattito pubblico. Meglio occuparsi, insomma, più prudentemente e più modestamente, del presente. E per quanto riguarda il presente occorre riconoscere, come ha evidenziato Paolo Benanti a proposito di ChatGPT – il LLM più noto e familiare anche ai profani –, «l’introduzione dell’apprendimento per rinforzo in un modello come ChatGPT ha effetti significativi sulla semantica delle risposte generate. Ottimizza ad esempio le risposte basate sul contesto, ma non trasforma il modello in un sistema dotato di versa comprensione semantica del senso umano»⁹.

⁵ Cfr. L. FLORIDI, *AI as Agency without Intelligence: On Artificial Intelligence as a New Form of Artificial Agency and the Multiple Realisability of Agency Thesis*, in *Philosophy & Technology*, 2025, art. n. 30. Cfr. anche la recente raccolta traduzione italiana ID., *La differenza fondamentale. Artificial Agency: una nuova filosofia dell’intelligenza artificiale*, Milano, 2025.

⁶ Cfr. J. SEARLE, *Is the Brain’s Mind a Computer Program?*, in *Scientific American*, 1, 1990, pp. 26-31. Sul punto cfr. A. ROMEO, *Testando il test dell’imitazione*, di prossima pubblicazione in *Diritto Artificiale*, 2, 2025.

⁷ Sulle diverse posizioni in relazione ai LLM, sinteticamente, cfr. A. SANTOSUOSSO, G. SARTOR, *Decidere con l’IA. Intelligenze artificiali e naturali nel diritto*, Bologna, 2024, pp. 55 s.

⁸ Cfr. *ivi*, p. 62.

⁹ P. BENANTI, *L’uomo è un algoritmo? Il senso dell’umano e l’intelligenza artificiale*, Roma, 2025, p. 36.

Dall'altro lato vi sono gli argomenti di chi avvicina e persino sovrappone l'IA e l'intelligenza umana sulla base di un riduzionismo computazionale della seconda. In estrema sintesi, si tratta degli argomenti di chi ritiene che, in fin dei conti, tutte le attività svolte dal cervello umano – ciò che complessivamente chiamiamo “intelligenza” – siano riducibili a elementi matematizzabili e associazioni sintattiche¹⁰. In questo modo, evidentemente, non si nega la natura soltanto sintattica dell'IA; al contrario, la si dà per assodata. Piuttosto, si nega l'esistenza di un'eccedenza semantica che sia esclusiva dell'intelligenza umana (o almeno la sua eventuale rilevanza). Ma se è così, l'area del disaccordo argomentativo non è più definita dal problema della natura dell'IA, bensì da quello della natura umana. In questo saggio, infatti, si intende sostenere che proprio nella riduzione sintattica operata dal GM, per cui, come è stato scritto, «l'individuo *non è altro che* un comportamento privo di significato che si può trattare come un segno»¹¹, si realizzi una grave lesione della dignità umana. Come ha recentemente sottolineato Benanti, proprio nella dimensione semantica che fa dell'essere umano un “essere di senso” si dà lo spazio della sua dignità: «Per ciascuno di noi, per la persona che siamo, la semantica apre una dimensione unica che è strettamente connessa a quanto esprimiamo con il concetto di dignità umana»¹².

Data la complessità del punto, tuttavia, prima di vedere in che modo l'utilizzo di IA sarebbe lesivo della dignità umana nel contesto della relazione tra il giudice e le parti nel processo, è indispensabile una riflessione sulla nozione di dignità umana.

3. La dimensione relazionale della dignità

L'impresa di dare una trattazione esaustiva della nozione di dignità umana non può aver luogo in questa sede. Nelle pagine che seguono si svolgerà una riflessione sintetica, limitata allo scopo di costruire il quadro concettuale essenziale al tema di questo saggio, legato eminentemente alla dignità nel suo aspetto relazionale. L'argomento procederà lungo i seguenti passaggi: *a)* in che senso le teorie della dotazione non possono fare a meno della dimensione relazionale; *b)* in che modo la teoria dello status non risolve i problemi sollevati dalla dimensione relazionale della dignità; *c)* in che senso la dignità informa normativamente la relazione tra il giudice e le parti nel processo.

a) Le teorie della dotazione e la dimensione relazionale

Nel dibattito sul concetto di dignità umana si parla di *teorie della relazione* per quelle teorie secondo cui la dignità si dà in quanto esito di una dinamica sociale intersoggettiva. La relazione, dunque, sarebbe costitutiva della dignità umana. Queste teorie si contrappongono alle *teorie della dotazione*, per le quali la dignità è una proprietà dell'essere umano in quanto tale. Si tratta delle cosiddette teorie ontologiche, nella cui categoria rientrano tanto la dottrina cattolica dell'uomo come *imago Dei*, quanto quella kantiana della natura razionale di ciascun individuo. Tuttavia, anche le teorie della dotazione devono riferirsi, in ultima istanza, a una

¹⁰ Per una critica di questa forma di riduzionismo, cfr. L. Floridi, A. C. NOBRE, *Anthropomorphising machines and computerising minds: the crosswiring of languages between Artificial Intelligence and Brain & Cognitive Sciences*, in *Minds and Machines*, 2024, art. n. 5 (anche questo argomento è sviluppato nella sopra citata traduzione italiana degli studi di Floridi). Ben più modestamente, sia pur concesso di rinviare a A. CAPUTO, *Not of Syntax Alone: AI and the Practice of Law*, in *Ordines*, 2, 2024, pp. 406-421.

¹¹ A. GARAPON, J. LASSÈGUE, *La giustizia digitale. Determinismo tecnologico e libertà*, Bologna, 2021, p. 218.

¹² P. BENANTI, *L'uomo è un algoritmo?*, cit., p. 43.

dimensione relazionale. È nella dinamica intersoggettiva, infatti, che si dispiega pragmaticamente il portato normativo della dignità in senso ontologico. Nel pensiero cattolico, la qualità di *imago Dei* dell'essere umano si riflette nel dettato evangelico per cui «tutto quello che avete fatto a uno dei più piccoli di questi miei fratelli, l'avete fatto a me»¹³. Nel sistema kantiano, la dignità dell'uomo in quanto agente morale autonomo fonda l'imperativo categorico di trattare l'altro sempre anche come fine e mai soltanto come mezzo.

Insomma, il fondamento ontologico non risparmia a queste teorie la prova della dimensione relazionale. Come ha scritto Francesco Viola, che ritiene una teoria della dotazione persino necessaria, «il carattere relazionale della dignità intesa in senso pratico o normativo implica che il riconoscimento sia un suo elemento costitutivo»¹⁴. Anche se la semplice appartenenza alla specie umana facesse di un individuo un soggetto dotato di dignità umana, al di fuori dell'incontro con l'altro questa condizione sarebbe del tutto inutile. Solo nella relazione si danno le pretese al rispetto della propria dignità e gli obblighi di riconoscimento della dignità altrui.

Sul piano pragmatico, insomma, anche intesa in questo senso la dignità umana è esposta alla problematicità delle relazioni umane: la *certezza* ontologica deve compromettersi con la multiforme *imprevedibilità* delle dinamiche intersoggettive.

b) La dimensione relazionale e la dignità come status

La relazione è concettualizzata in modi diversi nelle varie teorie: si parla di rispetto, cura, attenzione, riconoscimento. Ma pur nelle diverse sfumature, si tratta di varianti basate su un dato relazionale minimo. Al fondo della dinamica si dà sempre l'incontro tra un soggetto – *dignifier* – che tratta l'altro come degno – *dignified*. È chiaro che questo genere di relazione può darsi in un'infinita varietà di contesti informali di cui è impossibile fornire una casistica esaustiva¹⁵. Genericamente si può dire che le teorie *pure* della relazione, rinunciando a ogni fondamento ontologico, si trovano innanzi il problema di fondare la doverosità e il contenuto di questa relazione di *dignificazione*. In maniera diversa questo problema si pone anche per le teorie della dotazione: saldo il fondamento ontologico, bisogna chiedersi cosa di volta in volta questo dato imponga ai soggetti coinvolti nelle concrete dinamiche relazionali. È interessante, in questa sede, guardare al caso di quelle relazioni formalizzate in cui la dinamica intersoggettiva è codificata dal diritto. Si può ritenere, infatti, che le posizioni soggettive garantite dal diritto riassumano il contenuto della dignità e che, pertanto, per essere un *dignifier* basti rispettare il dovere giuridico di trattare l'altro, il *dignified*, secondo quanto stabilito dal diritto.

Una simile concezione della dignità come status sembra essere quella proposta da Jeremy Waldron. Secondo Waldron, la nozione di dignità va ricondotta all'antica idea di *rango*. Lungi dal segnare una cesura rispetto al passato di una società ordinata gerarchicamente, la moderna concezione di dignità umana sarebbe il risultato di un'universalizzazione dell'appartenenza al rango posto al vertice della piramide sociale. «As far as dignity is concerned», ha scritto Waldron, «the connotation of *ranking status* remained, and that what happened was that it was transvalued rather than superseded»¹⁶. Non v'è soluzione di continuità tra la dignità castale degli antichi e quella umana dei moderni: «The modern notion

¹³ Matteo 25, 40.

¹⁴ F. VIOLA, *Dignità umana, diritti e legge naturale*, in *Prudentia Iuris*, 2017, p. 39.

¹⁵ Cfr. F. SCAMARDELLA, *La dimensione relazionale come fondamento della dignità umana*, in *Rivista di filosofia del diritto*, 2, 2013, pp. 305-320.

¹⁶ J. WALDRON, *Dignity and Rank*, in ID., *Dignity, Rank, and Rights. The Tanner Lectures on Human Values*, Oxford, 2009, p. 227 (corsivo nel testo).

of *human* dignity involves an upwards equalization of rank, so that we now try to accord to every human being something of the dignity, rank, and expectation of respect that was formerly accorded to nobility»¹⁷. Una teoria come questa, in buona sostanza, tende a risolvere la dignità umana nel corredo di diritti forniti dal moderno istituto della cittadinanza che, appunto, generalizza per tutti i membri della comunità politica quella «upwards equalization of rank» che fa della società moderna un consesso di nobili.

Tale maniera di pensare la dignità umana presenta evidenti criticità. Come ha evidenziato Viola, «l'identificazione della dignità con il complesso di diritti che sono di fatto attribuiti nei contesti sociali si configura come un argomento circolare: abbiamo dignità perché abbiamo diritti inviolabili e viceversa. Con ciò stesso la dignità diventa una nozione inutile o reduplicativa»¹⁸. Il tentativo di fare a meno di un sostegno ontologico conduce Waldron a lasciare la nozione di dignità umana in balia delle contingenti determinazioni socioculturali che si manifestano nel diritto positivo. In tal modo, la dignità umana risulterebbe appiattita sullo sfondo normativo della moralità positiva di una certa comunità politica, codificata nel suo ordinamento giuridico. E ciò avrebbe effetti difficili da accettare. Si andrebbe, infatti, incontro al corto-circuito per cui dovrebbe ritenersi impossibile ledere la dignità umana mediante condotte rispettose del diritto positivo; ne uscirebbe pertanto neutralizzato il portato critico della dignità umana¹⁹. Priva di fondamenti extra-giuridici, in altre parole, la dignità umana perderebbe il suo potenziale normativo.

Waldron, invero, non è insensibile a questa problematica. Interrogandosi sull'esistenza di una connessione «less contingent than this»²⁰ tra dignità e diritto, egli giunge a sostenere che la dignità costituisce una «legal idea»²¹ non contingente: «law inherently promotes dignity»²². Waldron ancora allora questa posizione a una definizione minima di dignità: «Dignity is the status of a person predicated on the fact that she is recognised as having the ability to control and regulate her actions in accordance with her own apprehension of norms and reasons that apply to her»²³. Il diritto, secondo Waldron, consiste in un modo di esercitare il potere rispettando e promuovendo la dignità in questo senso. Egli sostiene che la maniera in cui si manifesta la dimensione coercitiva del diritto sia concettualmente dipendente da un'adesione interna a quest'idea di dignità. Tale dimensione, infatti, si dispiega per un verso nel senso della *self-application*, cioè in quella caratteristica delle norme per cui, non disciplinando esse stesse la propria applicazione, si lascia agli individui la responsabilità di farlo²⁴, e per altro verso nella dimensione dialogica del processo, attraverso cui si riconosce ai soggetti la capacità di presentare il proprio punto di vista sul diritto.

Potrebbe ritenersi che questo ricorso obbligato a una definizione che precede il diritto positivo dimostri l'indispensabilità, come ritiene Viola, di un fondamento ontologico per argomentare in termini di dignità umana. La definizione di Waldron, del resto, ricalca quella kantiana, benché letta in termini di status. La questione dei fondamenti, come si vedrà fra

¹⁷ *Ivi*, p. 229 (corsivo nel testo). Sulle tesi di Waldron, cfr. M. ATIENZA, *Sobre la dignidad humana*, Madrid, 2022, pp. 117 ss.

¹⁸ F. VIOLA, *Dignità umana*, cit., p. 42.

¹⁹ Cfr. *ibid.*

²⁰ J. WALDRON, *How Law Protects Dignity*, in *Cambridge Law Journal*, 1, 2012, p. 201.

²¹ *Ivi*, p. 203.

²² *Ibid.*

²³ *Ibid.*

²⁴ A proposito di IA, il tema della normatività si ricollega alle posizioni espresse da Roger Brownsword e William Lucy; sul punto, cfr. A. ROMEO, *La norma non può morire. Brevi riflessioni critiche sugli ultimi necrologi del diritto e sulla gestione tecnologica della condotta umana*, in *Diritto & Questioni Pubbliche*, 1, 2024, pp. 81-99.

poco, può anche essere lasciata in sospeso; sulla definizione di dignità si tornerà più avanti. Quel che qui interessa argomentare è che non sembra possibile ragionare efficacemente in termini di dignità umana se non si riconosce a questa nozione uno statuto semantico e normativo indipendente dal diritto positivo.

Waldron, del resto, ammette che questo status, vista l'impossibilità di prevedere i modi in cui viene esercitato il potere, potrebbe anche non coincidere con quello giuridicamente stabilito. In questo senso l'idea di dignità «has to function as a normative idea: it is the idea of a certain status that *ought to be* accredited to all persons and taken seriously in the way they are ruled»²⁵. Ma se è così, il rango non può essere concepito come una mera posizione giuridica soggettiva. Si tratta, piuttosto, di un orizzonte normativo cui il diritto, per essere autenticamente diritto, deve tendere. Lo stesso Waldron sostiene che un ordinamento giuridico non connesso all'idea di dignità non dovrebbe essere considerato «a true system of law»²⁶.

La connessione tra ordinamento giuridico e dignità, però, in questo modo rimane contingente. L'idea della dignità umana esercita una pressione normativa sul diritto, ma ciò non implica che le forme del diritto di per sé garantiscano che tale pressione sia necessariamente soddisfatta. È ben possibile che non vi sia un rimedio definitivo alla contingenza. Tuttavia, piuttosto che considerare, come fa Waldron, il rispetto della dignità umana un elemento necessario di un fenomeno contingente, parrebbe più opportuno riconoscere alla dignità umana un'eccedenza assiologico-normativa rispetto a quanto stabilito dal diritto. Il che significa che, oltre a non potere essere ridotta a un istituto giuridico, la dignità umana non può nemmeno essere considerata un aspetto necessario del concetto di diritto. Piuttosto, la dignità deve essere pensata come principio normativo che deve orientare la pratica del diritto. Il rispetto della dignità, in altre parole, rimane un dovere in capo a chi *usa* il diritto e non può mai dirsi assolto a monte dalle forme del diritto positivo.

L'idea dello status giuridico, pertanto, non elide la problematicità della dinamica relazionale. Che la relazione sia considerata costitutiva della dignità, oppure solo l'atto con cui la dignità altrui si riconosce e si rispetta, in ogni caso la dignità è in sospeso fino al suo dispiegamento pragmatico: in sospeso tra la pretesa del *would-be dignified* e il dovere dell'*ought-to-be dignifier*. Nelle relazioni codificate dal diritto un soggetto non è *dignified* dalla mera osservanza altrui delle norme vigenti: occorre anche che l'altro si incarichi di interpretare e applicare le norme *nel senso* imposto dal dovere di farsi *dignifier*. In altre parole, se è sostenibile una pretesa ad essere trattati degnamente questa non può risolversi in una semplice pretesa di legalità. Si tratta, piuttosto, della pretesa a un uso del diritto che non manchi di riconoscere e rispettare la propria dignità. A questo proposito, per l'argomento di questo saggio si deve guardare a quella particolare relazione giuridica che si configura tra il giudice e le parti in ambito processuale, invocata dallo stesso Waldron come caso paradigmatico della tutela giuridica della dignità umana.

c) *La dignità nella relazione tra il giudice e le parti*

Detto che il diritto positivo non esaurisce la questione della dignità, se non si vuole rinunciare a questa nozione bisogna rintracciarne i lineamenti fuori dal dettato giuridico. Come anticipato, Waldron ne offre una definizione minima sostanzialmente coincidente con la concezione kantiana, e individua nel processo un caso paradigmatico del modo in cui il diritto riconosce e promuove la dignità in questo senso. Le norme che regolano il processo,

²⁵ J. WALDRON, *How Law Protects Dignity*, cit., p. 203 (corsivo aggiunto).

²⁶ *Ivi*, p. 215.

secondo Waldron, sono costruite attorno al diritto delle parti di argomentare la propria posizione, e in tal modo ne riconoscono e rispettano l'autonomia morale. Come si è detto, tuttavia, l'esistenza di certe norme può certo esserne una condizione, ma di per sé non basta a garantire il rispetto e il riconoscimento della dignità; bisogna che il giudice le intenda e le usi in modo da generare questo effetto di riconoscimento e rispetto. In altre parole, le posizioni giuridiche delle parti e del giudice e la disciplina dei loro rapporti non risolvono in astratto la questione della dignità. La pretesa ad essere trattati degnamente, da un lato, e il dovere di riconoscere l'altro come degno restano in sospenso fino alla pratica del processo. Pertanto bisogna innanzitutto chiedersi in cosa consista, nel processo, la pretesa delle parti di essere trattati degnamente.

Si potrebbe rispondere, come fa Waldron, invocando il principio kantiano: essere trattati degnamente significa essere riconosciuti come moralmente autonomi. Tuttavia, questo approccio ontologico alla questione potrebbe non rappresentare la via migliore alla soluzione del problema. Scendere dal presupposto metafisico dell'ontologia classica per giungere a riempire di contenuto la pretesa individuale, infatti, pare una scelta argomentativa esposta a forti rischi, a cominciare da quello di una contestazione scettica dei fondamenti trascendentali. Qui non interessa confutare le teorie ontologiche. Ciò su cui si insiste è che, anche aderendo a una teoria della dotazione forte, il presupposto trascendentale della dignità umana risulta di per sé difficilmente operazionalizzabile nella concretezza della dimensione relazionale, se non mediante un lavoro di concretizzazione dagli esiti tutt'altro che scontati.

D'altro canto, la strategia argomentativa delle teorie della relazione potrebbe sollevare il problema opposto. Guardare alla concretezza delle diverse dinamiche intersoggettive, cioè, potrebbe fornire una nozione di dignità umana non universalizzabile e priva di capacità normativa. Tuttavia, queste teorie muovono da una tipizzazione astratta delle relazioni umane, distinguendo tra *relazioni di umiliazione* e *relazioni di dignificazione*. L'altro è umiliato se trattato al di sotto di quel che imporrebbe la sua qualità di essere umano; è, invece, dignificato se trattato conformemente ad essa. In questo presupposto, si avverte poco la differenza con le teorie della dotazione. La differenza si dà nel fatto che piuttosto che individuare in via trascendentale una qualità umana che definisca la dignità, da questa prospettiva si guarda alle diverse proprietà umane che vengono in rilievo nelle dinamiche intersoggettive. È anche possibile che muovendo da un presupposto ontologico come quello kantiano si giunga alle stesse conclusioni. Cambia, però, la domanda. Nel caso in esame non ci si chiede *cosa impone il rispetto della dignità umana nella relazione tra il giudice e le parti?* bensì *quale proprietà umana delle parti il giudice deve rispettare per costruire con esse una relazione dignificante?*

A sviluppare questo argomento è David Luban. Nel processo, secondo Luban, umiliare significa «treating a person's story and viewpoint as insignificant»²⁷. Secondo quanto appena detto, perché ciò costituisca una lesione della dignità umana questo aspetto deve essere ancorato a una proprietà dell'essere umano. Per far ciò, Luban non vuole far ricorso alla metafisica. Tuttavia, invece che provare a sbarazzarsi del tutto dell'ontologia, egli sviluppa il suo argomento a partire da una sorta di ontologia minima. Piuttosto che muovere da un presupposto trascendentale, egli prende le mosse dall'esistenza individuale, dalla *soggettività*. Per soggettività Luban intende quell'*ontological heft* per cui «each of us is an "I"»²⁸. Si tratta del punto attorno a cui ruota il mondo di ciascuno. Questa posizione dell'*io* al centro del cosmo, si badi, non ha alcuna implicazione assiologica o normativa. Utilizzando la terminologia di

²⁷ D. LUBAN, *Lawyers as Upholders of Human Dignity (When They Aren't Busy Assaulting It)*, in *University of Illinois Law Review*, 2005, p. 822.

²⁸ *Ivi*, p. 821.

Ralph B. Perry, Luban chiama questa condizione *egocentric predicament*²⁹, ma senza alcuna connotazione morale. Si tratta soltanto del dato esistenziale per cui ciascuno ha sé stesso come punto di vista sulla realtà: per ogni essere umano il mondo esiste in quanto passa attraverso la propria soggettività.

L'argomento di Luban, dunque, inizia dalla prima persona singolare. Ciascun essere umano è una storia vissuta dall'*io*. Ma finché si rimane in una dimensione solitaria e autoriferita, la soggettività non ha alcuna rilevanza in termini di dignità. Anche se il solo fatto di essere una storia fosse sufficiente a fare di questo *io* isolato una creatura dotata di dignità umana, ciò non avrebbe alcun reale significato. La questione della dignità emerge soltanto quando si configura la seconda persona, il *tu*, quando nell'incontro con l'altro la propria storia diventa oggetto di un racconto. Quando al proprio racconto non viene attribuita rilevanza, la negazione della propria qualità di *io*, del proprio *ontological heft*, fa di questa dinamica intersoggettiva una relazione di umiliazione. Si è dignificati, invece, quando la propria storia è tenuta in debito conto: essere trattati come esseri umani, in altre parole, significa che la propria storia viene ascoltata e compresa. La pretesa ad essere trattati degnamente così come la doverosità di trattare l'altro degnamente si fondano su una generalizzazione dell'*egocentric predicament*³⁰. Riconoscere l'altro come degno significa riconoscergli quel che sappiamo di noi stessi, cioè riflettere sull'altro la nostra esperienza della soggettività³¹. Si tratta di prendere sul serio l'altrui *ontological heft*: riconoscere agli altri *io* la posizione di centralità che nella soggettività attribuiamo a noi stessi.

Per costruirsi una relazione di dignificazione di questo genere, allora, non è sufficiente possedere la qualità di essere umano. Poiché la qualità di essere umano è riconosciuta all'altro sulla base della propria, è necessario anche che la relazione si dia tra esseri umani: nell'incontro tra *ontological hefts*, la reciprocità è costitutiva. Quando si dice "prendere sul serio" la storia altrui non ci si riferisce al dovere di registrare le informazioni che vengono trasmesse. Più profondamente, si intende quello sforzo di comprensione che passa per una generalizzazione dell'*egocentric predicament*: l'assunzione del punto di vista altrui come fosse il proprio.

Perché ciò avvenga nella relazione tra le parti e il giudice, Luban insiste sul ruolo svolto dall'avvocato³², ma questo aspetto non può essere affrontato in questa sede. Sarà sufficiente evidenziare che la tesi di Luban consente di individuare in maniera piuttosto soddisfacente, e sufficientemente universalizzabile, il contenuto della pretesa di dignità avanzata dalle parti nella relazione con il giudice: si tratta della pretesa di essere ascoltati e compresi quando si racconta la propria storia. Si può ora passare a vedere perché questa pretesa cadrebbe nel vuoto se le parti si trovasse di fronte un GM.

4. Giudice-macchina e dignità umana: Error 404

Secondo quanto detto sinora: la mera esistenza delle norme che regolano il processo in forma dialogica non è sufficiente a garantire che la relazione tra il giudice e le parti sia una relazione di dignificazione; a questo fine è essenziale che il giudice interpreti e applichi tali

²⁹ Cfr. *ibid.*

³⁰ Cfr. *ibid.*

³¹ Cfr. anche A. GEWIRTH, *Dignity as the Basis of Rights*, in M. J. MEYER, W. A. PARENT (eds.), *The Constitution of Rights. Human Dignity and American Values*, Ithaca-London, 1992, pp. 10-28.

³² Cfr. D. LUBAN, *The Rule of Law and Human Dignity: Re-examining Fuller's Canons*, in *Hague Journal on the Rule of Law*, 2, 2010, p. 41 ss.

norme assumendosi la responsabilità di farsi *dignifier* rispetto alle parti. Si è aggiunto, in conclusione, che per far ciò il giudice deve essere un giudice umano. Nella parte conclusiva del saggio ci si concentrerà su quest'ultimo aspetto. Quel che si vuole sostenere è che il modo stesso in cui l'IA funziona impedisce la possibilità che un GM possa agire da *dignifier* rispetto alle parti.

Nei dibattiti sulla prospettiva del GM si argomenta spesso a partire da un confronto con il giudice umano in termini di performance. Un GM, a differenza di quello umano, sarebbe instancabile, non soggetto a distrazioni o dimenticanze, rapido e infallibile nelle ricerche e nei calcoli. Oltracciò, un GM sarebbe immune non soltanto alle debolezze, per così dire, strutturali dell'essere umano, ma anche a quelle morali: non conoscerebbe la crudeltà, l'avidità, il rancore, la paura, ma nemmeno l'ebbrezza conferita dall'esercizio del potere, le ambizioni di carriera, i condizionamenti ambientali. Quanto all'argomento ormai classico della discriminazione algoritmica, si ha gioco facile nel sostenere che anche il ragionamento umano è fortemente esposto al rischio che *bias* e pregiudizi di vario genere ne influenzino l'andamento e il risultato. Ci si può chiedere: tutto ciò è davvero efficacemente compensato da quel tratto tipicamente umano, l'empatia, che spesso si invoca a difesa del giudice umano? Per dare una risposta adeguata pare necessario porsi in un ordine di idee diverso da quello di una *gara di difettosità* tra l'uomo e l'IA. A questo fine, l'argomento procederà come segue: innanzitutto si vedrà in che senso esiste nel processo uno spazio in cui una caratteristica simile è necessariamente rilevante; in secondo luogo si affronterà il significato dell'empatia per fornirne una definizione non romantica; infine si sosterrà che lasciare all'IA il compito di svolgere questa funzione impedirebbe il riconoscimento della dignità umana delle parti.

Quando Beccaria parlava della decisione giudiziale come del risultato di un sillogismo perfetto, e Montesquieu invocava la figura del giudice *bouche de la loi*, l'illuminismo giuridico sognava qualcosa di simile a una concezione algoritmica del ragionamento giudiziario. Se la giurisprudenza si svolgesse nel modo perfettamente meccanico delle forme logiche del sillogismo, un'IA ben programmata potrebbe soddisfare in pieno le aspettative illuministe. Tuttavia, la prospettiva di questa perfezione logica è nel dibattito filosofico-giuridico ampiamente contestata. In questa sede si può tralasciare il tema della determinazione della premessa maggiore del sillogismo mediante interpretazione dei testi normativi. Non perché tale dimensione del giudizio non sollevi particolari preoccupazioni nel caso del GM, ma soltanto perché affrontarle porterebbe lontano dall'oggetto di questo studio. Si supporrà, pertanto, che il GM sia in grado di svolgere in questa dimensione del giudizio pressoché la medesima funzione del giudice umano senza che ciò incida sul senso del processo.

Si guardi, invece, alla verifica fattuale dell'oggetto del giudizio e al procedimento che conduce alla decisione. Secondo il modello sillogistico, a una precisa ricostruzione dei fatti dovrebbe seguire automaticamente l'applicazione della norma. Nella realtà, tuttavia, avviene qualcosa di ben diverso. Tra la verifica fattuale e la decisione si interpone quella che Luigi Ferrajoli ha definito *terza dimensione del giudizio*: la dimensione equitativa. Si utilizza qui la posizione di Ferrajoli perché è particolarmente importante mettere in evidenza che anche gli studiosi più critici della cosiddetta *creatività giurisprudenziale* riconoscono l'ineludibilità di questa dimensione. Qui non rilevano le controversie teoriche che mettono distanza tra Ferrajoli e chi considera la dimensione equitativa uno spazio per la creatività giurisprudenziale. Più interessante delle divergenze è qui l'aspetto della convergenza delle diverse teorie del ragionamento giudiziale sulla funzione specifica di questa dimensione. È grazie all'argomentazione equitativa che entrano nel procedimento che conduce alla decisione gli aspetti particolari e concreti del fatto. Con le parole di Ferrajoli, l'argomentazione equitativa implica «la comprensione e il bilanciamento delle

circostanze irripetibili che fanno di ciascun caso un caso differente dagli altri, pur se tutti sussumibili nella stessa fattispecie legale»³³.

L'argomentazione equitativa segue un certo modo di ricostruire il fatto che non si limita a un processo cognitivo. Tale ricostruzione implica non solo un'elencazione degli elementi particolari del fatto, ma anche uno sforzo di comprensione mediante il quale il giudice attribuisce un certo significato a questi elementi, un certo peso, e li mette in connessione. È qui che viene in rilievo l'indispensabilità di una capacità specificamente umana che l'IA non può replicare. Quel che si chiede al giudice è di prendere sul serio le storie raccontate dalle parti, il loro punto di vista sul mondo. Ciò significa che la comprensione deve svolgersi attraverso un riconoscimento dell'altrui soggettività come *egocentric predicament*. Quando una parte racconta la sua storia non sta solo elencando fatti; sta portando all'attenzione del giudice tutto l'ordito di convinzioni, esperienze, relazioni, speranze ed errori che passa per il suo *io*. Comprendere la sua storia significa assumere la validità ontologica di questo punto di vista. Significa, come ha scritto Massimo Lo Giudice in un recente libro che ben mette in luce la natura irriducibilmente umana del processo, *mettersi al posto degli altri*³⁴. Quando si parla, spesso in maniera molto vaga, di empatia, sembra in fin dei conti farsi riferimento a una capacità di questo genere, che qui si è definita in termini privi di ogni accesso sentimentalistico, come la capacità di guardare alla storia altrui *dal di dentro*.

E solo in questo modo si può guardare ai fatti senza che ciò anticipi l'atto del giudicare. Perché, come scrive ancora Ferrajoli, entrino nel giudizio «il contesto ambientale del fatto, le condizioni in cui esso è stato vissuto, le effettive circostanze umane e sociali nelle quali risiede la sua concretezza e particolarità e che lo rendono diverso da tutti gli altri»³⁵, bisogna che a questi elementi sia riconosciuta la loro unicità e irripetibilità, che si dà soltanto in connessione con la soggettività delle parti. Solo in questo modo si riconosce all'individuo quella proprietà umana, essere un *io*, che coincide con la sua dignità: all'*io* del giudice parla l'*io* delle parti. Ed è in questo che l'IA non può sostituirsi senza ledere la dignità umana.

Come sopra detto, l'IA non è capace di comprensione. Il suo funzionamento è vincolato all'associazione probabilistica tra elementi sintattici e non ha accesso alla dimensione semantica. Ciò implica che, per un GM, prendere in considerazione il racconto delle parti equivarrebbe a un processo di conversione dei fatti in dati mediante scomposizione in elementi sintattici, cioè traduzione in elementi calcolabili. Per essere letti dalla macchina, i fatti umani devono essere trattati come 0 e 1. Così, disponendo di elementi omologhi alla mole spaventosa di dati di cui *già* dispone, essa può svolgere un'imitazione dell'atto di comprensione. Il punto di vista dell'*io* si perde totalmente nell'attribuzione di significato alla storia, che invece avviene, nella macchina, associando probabilisticamente i dati raccolti nel contesto su un piano di assoluta commensurabilità e fungibilità con i dati preesistenti, estranei alla vicenda oggetto del giudizio. L'*io* si perde in una forzata omologia col mondo.

Il principio del mondo stabile agisce per l'IA, pur in senso probabilistico, come meta-criterio di verità. Il significato e il peso dei fatti sono stabiliti per mezzo di connessione basate su questo principio, e così le conclusioni che portano al giudizio³⁶. A meno che non si suppone la possibilità che il racconto delle parti si manifesti in una coincidenza perfetta tra la dimensione semantica e quella sintattica, l'argomentazione equitativa, operata dal GM, è svuotata del suo senso. La supposizione poggia su un'ipotesi palesemente insostenibile, ma individua in buona sostanza il principio di funzionamento dell'IA. Come è stato scritto: «l'ideologia digitale assimila

³³ L. FERRAJOLI, *La costruzione della democrazia. Teoria del garantismo costituzionale*, Roma-Bari, 2021, p. 159.

³⁴ Cfr. M. LO GIUDICE, *Il dramma del giudizio*, Milano-Udine, 2023, p. 181 s.

³⁵ L. FERRAJOLI, *Diritto e ragione. Teoria del garantismo penale*³, Roma-Bari, 1996, p. 145.

³⁶ Cfr. A. SANTOSUOSSO, G. SARTOR, *Decidere con l'IA*, cit., pp. 138 ss.

[...] il *socius* a una oggettività dello stesso ordine di quella che si trova in natura»³⁷. E ciò comporta una mortificazione della soggettività nell'oggettività algoritmica: l'individuo è così *homo numericus* e non *homo dignus*³⁸.

Per concludere questo saggio è opportuno segnalare brevemente due criticità. Innanzitutto, si potrebbe ritenere che la capacità dell'IA di simulare la comprensione renda la relazione tra le parti e un GM non dissimile da una relazione tra le prime e un giudice umano. Riprendendo le classiche tesi di Turing, una posizione simile potrebbe argomentarsi a partire dal fatto che, nell'interazione umana, al soggetto non è data la facoltà di conoscere realmente lo stato psichico altrui. Non potendo accedere all'intimità dell'altro, il soggetto deve accontentarsi del dato esteriore offerto dal suo interlocutore, e supporre il resto. Ma ciò rende la simulazione algoritmica della comprensione e l'interazione umana fenomenologicamente identiche. Non ci si riferisce, beninteso, all'eventualità che l'IA possa fornire una risposta giusta o coincidente con quella offerta da un giudice umano. Il punto è che questa equivalenza fenomenologica potrebbe fondare l'argomento per un'equivalenza qualitativa. Il complesso tema è oggetto di un recente studio assai articolato di Andrea Romeo di cui si può qui solo accennare l'argomento³⁹. Secondo Romeo, tale equivalenza non è sostenibile in virtù della natura funzionale del linguaggio dell'IA. Giacché è costruito su un programma di simulazione, il linguaggio funzionale è orientato allo scopo pratico della comunicazione, ma fallisce del tutto sul piano performativo. Nei suoi enunciati non c'è alcuna pretesa di correttezza: anche a costo di contraddirsi, il linguaggio funzionale giustifica in maniera circolare i suoi enunciati con l'esecuzione del compito di produrre enunciati dotati di senso. Al contrario, l'interazione umana si basa sulla possibilità di chiedere le ragioni che fondano il dato esteriore degli atti linguistici, giacché essi poggiano su una pretesa di correttezza e non su un aspetto funzionale. Le parti, insomma, non possono conoscere l'animo del giudice, ma di fronte alle sue decisioni possono avanzare una fondata pretesa di giustizia.

In secondo luogo, si potrebbe ritenere che l'argomento proposto, rigettando il giudizio automatico prodotto dall'oggettività algoritmica, conduca per le vie della soggettività a un giudizio impossibile. L'enfasi sulla soggettività parrebbe infatti negare quella distanza tra il giudice e le parti che non solo è garanzia di imparzialità, ma altresì evita che si attivi il corto-circuito per cui *comprendere tutto è perdonare tutto*. La questione attiene a una dimensione del giudizio che esula dal perimetro tematico di questo saggio. Tuttavia, vale la pena segnalare almeno che non bisogna fare l'errore di credere che la comprensione sostituisca il giudizio. Attraverso il riconoscimento dell'altrui soggettività, la comprensione prepara un giudizio che non può seguire le forme dell'automatismo algoritmico. Essa fornisce al giudice gli elementi per esercitare, con l'argomentazione equitativa, quella facoltà con cui l'*io* interagisce con la complessità del mondo: il buon senso. Si tratta, in estrema sintesi, della *phronesis* aristotelica⁴⁰, di quel «surrogato pragmatico del senso del tutto»⁴¹, come ha scritto Lo Giudice, che Ferrajoli definisce *iuris-prudentia*⁴². È questa capacità a distinguere la funzione giudiziaria «da qualunque altra funzione burocratica o meramente esecutiva»⁴³, e la decisione umana da quella robotica⁴⁴.

³⁷ A. GARAPON, J. LASSÈGUE, *La giustizia digitale*, cit., p. 106.

³⁸ Cfr. S. RODOTÀ, *Il diritto di avere diritti*, Roma-Bari, 2015, p. 197.

³⁹ Cfr. ancora A. ROMEO, *Testando il test dell'imitazione*, cit., di prossima pubblicazione.

⁴⁰ Cfr. A. GARAPON, J. LASSÈGUE, *La giustizia digitale*, cit., p. 275.

⁴¹ M. LO GIUDICE, *Il dramma del giudizio*, cit., p. 171.

⁴² Cfr. L. FERRAJOLI, *Diritto e ragione*, cit., p. 145.

⁴³ *Ibid.*

⁴⁴ Cfr. G. GIGERENZER, *Perché l'intelligenza umana*, cit., p. 120 s.



MARCO SILIGATO*

IL SETTORE DELLA GIUSTIZIA NELL'AI ACT: IL DOVERE DI TRASPARENZA COME “ANIDRIDE SOLFOROSA”?

SOMMARIO: 1. Introduzione. – 2. Profili strutturali dell'AI Act e applicazione al settore della giustizia. – 3. L'utilizzo dei sistemi IA nel settore della giustizia penale: fra attività vietate e ad «alto rischio». – 4. Il dovere di trasparenza: reale garanzia?. – 5. Conclusioni.

1. Introduzione

L'idea al centro del presente lavoro affonda le radici nella potente metafora dell'anidride solforosa: con tale immagine, Lucio Dalla, già nel 1975 riportava in versi i propri dubbi in merito al futuro uso dei sistemi informatici, ai rischi legati all'uso di uno strumento calcolatore alienante e disumanizzante, anticipando profeticamente l'impatto dirompente e il potenziale effetto degli strumenti informatici sulla società, ben prima che la nuova tecnologia entrasse nel dibattito sociale.

Traendo spunto da tale provocazione, seppur non ispirata all'intelligenza artificiale, si indagherà sull'attuale regolamentazione in merito alla tutela dei diritti fondamentali riconosciuta agli individui, per provare a comprendere se l'accesso al nuovo strumento informatico sia ancora “fumoso”.

Le tecnologie più moderne basate su sistemi di intelligenza artificiale (IA) hanno ormai fatto ingresso in ogni ambito della vita quotidiana nonostante tale strumento sia talmente nuovo che non se ne conoscono ancora la reale portata e dimensione.

La nuova strumentazione informatica, sulla base di una sempre maggiore disponibilità di dati legati a nuovi canali digitali e spazi comunicativi *online*, ha infatti dato vita a quello che è stato definito un “ecosistema digitale”¹, offrendo supporto ai cittadini tramite riferimenti e approfondimenti, anche in ambito giudiziario.

* Dottorando di ricerca in Studi europei, Dottorato di Interesse Nazionale (DIN) dell'Università di Genova, *Curriculum Governance multilivello e diritti fondamentali*, sede di afferenza Università degli Studi di Messina, Dipartimento di Scienze Politiche e Giuridiche.

¹ A. CORRERA, *Il ruolo dell'intelligenza artificiale nel paradigma europeo dell'E-justice. Prime riflessioni alla luce dell'AI Act*, in *Quaderni AISDUE*, n. 2/2024, p. 2.

In tale ottica, il presente lavoro intende analizzare il «dovere di trasparenza» così come delineato nel Regolamento (UE) n. 1689/2024 (d'ora innanzi "*AI Act*"), sottolineandone la specificità e le implicazioni pratiche, nonché i limiti che possono emergere da definizioni non ancora applicate e interpretate.

Dopo aver presentato, seppur sinteticamente, i contenuti essenziali della regolamentazione del settore della giustizia penale dell'*AI Act*, evidenziando le peculiarità applicative degli strumenti di IA in tale ambito, si procederà all'analisi del dovere di trasparenza così come enucleato nel Regolamento, esaminando la coerenza e l'effettività di tali previsioni, nonché i profili di spiegabilità e responsabilità.

In questo senso, l'analisi dei riferimenti normativi dell'*AI Act* verterà sull'imposizione ai vari soggetti coinvolti (utilizzatori, fornitori, ecc.), di divieti o obblighi tecnico-procedurali a tutela della trasparenza, graduati in base al settore di applicazione e allo *standard* di tutela richiesto, mediante una "procedimentalizzazione" del rischio³, con particolare riguardo al settore della giustizia penale.

Nelle conclusioni, dunque, volendo sfuggire dalla mera ritualità, partendo dal citato ambito applicativo dei sistemi di IA nel settore della giustizia penale, si tenterà di fornire un punto di vista circa la valutazione dottrinale del livello di garanzia della disciplina relativa a tale settore, nell'ambito della recente regolamentazione europea.

2. Profili strutturali dell'*AI Act* e applicazione al settore della giustizia

La rapida diffusione dei sistemi basati sull'IA ha generato un dibattito globale sulla necessità di un quadro normativo capace di bilanciare innovazione tecnologica e tutela dei diritti fondamentali.

In questo quadro si inserisce il solido approccio europeo adottato dalla Commissione europea già nel 2020 con il Libro bianco sull'intelligenza artificiale⁴ e sviluppatosi negli anni attraverso numerosi atti legislativi di varia forma⁵, di recente cristallizzatosi nell'*AI Act*, con

² Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024, *che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale)*, in GUUE Serie L del 12/7/2024.

³ A. FORMISANO, *L'impatto dell'intelligenza artificiale in ambito giudiziario sui diritti fondamentali*, in *Federalismi.it*, n. 22/2024, p. 136.

⁴ Commissione europea, *Libro bianco sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia*, del 19 febbraio 2020, COM(2020) 65 final.

⁵ Commissione europea, *Piano coordinato sull'intelligenza artificiale*, del 7 dicembre 2018, COM(2018) 795 final (modificato il 21 aprile 2021); Id., *Creare fiducia nell'intelligenza artificiale antropocentrica*, dell'8 aprile 2019, COM/2019/168 final; Id., *Libro bianco sull'intelligenza artificiale Un approccio europeo all'eccellenza e alla fiducia*, del 19 febbraio 2020, COM(2020) 65 final; Id., *Relazione sulle implicazioni dell'intelligenza artificiale, dell'Internet delle cose e della robotica in materia di sicurezza e di responsabilità*, 19 febbraio 2020, COM (2020) 64 final; Parlamento europeo, *Risoluzione sui diritti di proprietà intellettuale per lo sviluppo di tecnologie di intelligenza artificiale*, del 20 ottobre 2020, (2020/2015(INI)); Id., *Risoluzione sui processi decisionali automatizzati: garantire la tutela dei consumatori e la libera circolazione di beni e servizi*, del 12 febbraio 2020, (2019/2915(RSP)); Id., *Risoluzione sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale*; *Dichiarazione europea sui diritti e i principi digitali per il decennio digitale*, 2023/C 23/01, del 6 ottobre 2021, P9_TA(2021)0405; Corte dei Conti UE, *Relazione speciale 08/2024: Le ambizioni dell'UE in materia di intelligenza artificiale – Per il futuro, una governance più forte e investimenti più consistenti e mirati sono essenziali*, del 29 maggio 2024.

cui l'Unione, perseguendo l'ambizioso obiettivo di guidare lo sviluppo dell'IA a livello mondiale⁶, ha istituito vincoli uniformi e direttamente applicabili all'interno degli Stati membri in materia di utilizzo di sistemi di intelligenza artificiale⁷.

Ai fini del presente lavoro pare opportuno, seppur brevemente, riferire in merito alla duplice base giuridica del Regolamento⁸: l'articolo 114 TFUE, in virtù del quale l'Unione può adottare misure volte al ravvicinamento delle normative nazionali in tema di funzionamento del mercato interno, e l'articolo 16 TFUE, sulla base del quale l'Unione può adottare atti legislativi in materia di protezione dei dati personali.

In tale ottica, infatti, si inserisce la finalità di promozione dell'innovazione e di tutela dei diritti fondamentali dell'*AI Act*⁹, coerente con la scelta di legare la pervasività degli interventi in base al grado di rischio di ciascun sistema secondo il c.d. *risk-based approach*¹⁰.

Entrando nel dettaglio della struttura dell'*AI Act*, il Regolamento distingue quattro tipi di sistemi IA in base al livello di rischio riconosciuto, definiti quali: inaccettabile, alto, moderato e basso¹¹.

Come già accennato, per ciascun livello il Regolamento ha graduato la portata della disciplina: per i sistemi a rischio inaccettabile è previsto il divieto di utilizzo dei sistemi IA con specifiche eccezioni¹²; per i sistemi a rischio alto, aprioristicamente ammissibili, sono previsti i necessari requisiti per l'applicabilità e l'utilizzo¹³; per i sistemi a rischio limitato sono previsti puntuali obblighi di trasparenza¹⁴; per quelli a rischio basso, la semplice sottoscrizione di codici di condotta¹⁵.

A tale quadro si aggiunge la figura dei cosiddetti modelli di IA per finalità generali (GPAI)¹⁶, così definiti in quanto non aprioristicamente classificabili in virtù della loro capacità di sviluppare abilità diverse rispetto a quelle predeterminate al momento della programmazione¹⁷.

Entrando nel merito dell'analisi di cui al presente paragrafo, deve sottolinearsi che l'Unione europea, dinanzi alla sfida posta dalla “nuova” intelligenza artificiale, ha rivolto

⁶ O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI, *La disciplina dell'intelligenza artificiale*, Milano, 2025.

⁷ P. INTURRI, S. FICHERA, A. COSTA, *La disciplina dei sistemi di intelligenza artificiale per l'amministrazione della giustizia nel Regolamento (UE) 2024/1689*, in *LavoroDirittiEuropa*, n. 1/2025, p. 2.

⁸ Per approfondimenti, si consiglia: E. CIRONE, *L'AI Act e l'obiettivo (mancato?) di promuovere uno standard globale per la tutela dei diritti fondamentali*, in *Quaderni AISDUE*, n. 2/2024, p. 4.

⁹ Regolamento (UE) 2024/1689, *cit.*, p.1; al considerando n. 1 si legge: «[Lo scopo del presente regolamento è] migliorare il funzionamento del mercato interno istituendo un quadro giuridico uniforme in particolare per quanto riguarda lo sviluppo, l'immissione sul mercato, la messa in servizio e l'uso di sistemi di IA in conformità con i valori dell'Unione, promuovere la diffusione di una IA antropocentrica e affidabile, garantendo nel contempo un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea, compresi la democrazia, lo Stato di diritto e la protezione dell'ambiente, proteggere contro gli effetti nocivi dei sistemi di IA nell'Unione, nonché promuovere l'innovazione».

¹⁰ G. ZICCARDI, *Una lettura dell'Artificial Intelligence Act: norme, etica, adempimenti, attuazione*, in AA. VV. (a cura di), *Intelligenza Artificiale. Diritto, giustizia, economia ed etica*, Pubblicazioni del Dipartimento di Scienze Giuridiche “Cesare Beccaria”, Bari, 2025, p. 20.

¹¹ S. QUATTROCOLO, *Intelligenza artificiale e processo penale: le novità dell'AI Act*, in *Diritto di difesa – La rivista dell'Unione delle Camere Penali Italiane*, 2025, p. 4.

¹² Regolamento (UE) 1689/2024, *cit.*, Capo II.

¹³ Regolamento (UE) 1689/2024, *cit.*, Capo III.

¹⁴ Regolamento (UE) 1689/2024, *cit.*, Capo IV.

¹⁵ Regolamento (UE) 1689/2024, *cit.*, Capo X.

¹⁶ Regolamento (UE) 1689/2024, *cit.*, Capo V.

¹⁷ Regolamento (UE) 1689/2024, *cit.*, art. 3, par. 1, n. 63.

l'attenzione anche ai sistemi giudiziari degli Stati membri¹⁸, comprendendo la necessità di compiere uno sforzo di adeguamento ulteriore per garantire un accesso alla giustizia moderno, efficiente, non invasivo e senza barriere, che rispetti il dogma dello Stato di diritto e che rafforzi la fiducia nel sistema istituzionale¹⁹.

Invero, dal considerando n. 61 del Regolamento si ricava la collocazione dei sistemi di intelligenza artificiale applicati alla giustizia tra le attività ad «alto rischio», precisandosi che «al fine di far fronte ai rischi di potenziali distorsioni, errori e opacità» sono da considerarsi tali i sistemi di intelligenza artificiale destinati a «essere utilizzati da un'autorità giudiziaria o per suo conto per assistere le autorità [...] nelle attività di ricerca e interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti».

Tale locuzione è infatti ripresa nell'Allegato III del Regolamento²⁰, quale unico caso di utilizzo delle applicazioni IA con impatto potenziale sulla sicurezza dei cittadini e sui diritti fondamentali nel settore della giustizia²¹.

In questo senso, circa la programmazione dei modelli di IA nei settori ad «alto rischio», l'AI Act precisa i criteri progettuali che devono garantire l'accessibilità, la qualità e l'adeguatezza dei *dataset* necessari e, allo stesso tempo, individuare, prevenire e colmare possibili distorsioni o gravi inadeguatezze nelle informazioni offerte agli utenti²².

Diventa dunque importante facilitare l'accesso ai dati che vengono utilizzati per «addestrare» i sistemi di IA al fine di poterne verificare la genuinità e la qualità, nonché per garantire la trasparenza della decisione assunta in base all'algoritmo.

Il principio, infatti, così come delineato nel Regolamento, sembra richiamare la più classica declinazione del principio di trasparenza in materia di diritto dei consumatori, da intendersi quale accesso continuativo ad informazioni intelligibili, assenza di ambiguità e arbitrarietà nei comportamenti assunti dal professionista, il quale è soggetto a determinati obblighi professionali²³.

Inoltre, nell'ambito del dibattito dottrinale sull'utilizzo dei sistemi di IA nel settore della giustizia, si è comunque concordi sull'utilizzo della strumentazione informatica quale strumento complementare all'intelligenza e alla sensibilità umana, soprattutto per la necessità di garantire il rispetto dei valori democratici su cui si fonda l'Unione europea²⁴.

3. L'utilizzo dei sistemi IA nel settore della giustizia penale: fra attività vietate e ad «alto rischio»

Preliminare all'analisi della relazione tra intelligenza artificiale (IA) e il settore della giustizia penale è una breve ma essenziale ricognizione del concetto di intelligenza artificiale come definito nel contesto della normativa dell'Unione Europea.

¹⁸ P. P. PAULESU, *Intelligenza artificiale e giustizia penale. Una lettura attraverso i principi*, in *Archivio penale*, n. 1/2022, p. 13.

¹⁹ A. CORRERA, *op. cit.*, p. 2.

²⁰ Regolamento (UE) 2024/1689, *cit.*, Allegato III, p. 127 e s.

²¹ Per un approfondimento, si veda O. CARAMASCHI, *Il costituzionalismo al cospetto dell'intelligenza artificiale: nuove sfide, quali soluzioni?*, in *Rivista italiana di informatica e diritto*, n. 1/2025.

²² *Ivi*, art. 10, par. da 2 a 5.

²³ F. DEANA, *La mano invisibile dell'Intelligenza Artificiale e il principio di trasparenza nei rapporti B2C: la tutela del consumatore nel mercato unico digitale*, in *Freedom, Security & Justice*, n. 1/2025, pp. 66-85.

²⁴ V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in AA.VV., *Intelligenza Artificiale – Il diritto, i diritti, l'etica*, U. RUFFOLO (a cura di), Milano, 2020.

Nello specifico, il concetto di «sistema IA» di cui all'articolo 3 dell'*AI Act*²⁵ è legato al riferimento a «livelli di autonomia variabile», dovendo pertanto espungersi dal nucleo concettuale i sistemi «automatizzati» che eseguono comandi senza alcuna forma di rielaborazione²⁶.

La «rivoluzione algoritmica» paventata in dottrina rappresenta dunque la nuova frontiera tecnologica in cui l'umano è chiamato a «misurarsi» con la macchina²⁷.

Merita certamente sottolineatura il mancato riferimento espresso alla natura «collaborativa» dell'intelligenza artificiale, potendo la stessa, secondo la lettera della norma, generare «decisioni».

Come osservato²⁸, tale riferimento non permette di escludere aprioristicamente la previsione di un modello di intelligenza artificiale «forte» volta «all'automazione del processo decisionale», in luogo di quella «debole», che collabora con la figura preposta in assenza di qualsivoglia autorità decisionale.

In questo senso, si è sostenuto che in un'ottica mitigatoria di tale accezione devono intendersi i dettami dei considerando n. 2 e n. 27, oltre che lo stesso articolo 1 del Regolamento, i quali precisano che l'intelligenza artificiale deve rispettare i criteri di antropocentrismo e affidabilità, garantendo così il necessario apporto umano²⁹.

A parere di chi scrive, debbono ritenersi coerenti anche le specifiche disposizioni relative ai settori ad «alto rischio» relative all'obbligo di garantire la trasparenza³⁰, la supervisione e la sorveglianza umana³¹, oltre a una valutazione sui possibili pregiudizi sui diritti fondamentali³².

Innanzitutto, deve segnalarsi che il Regolamento include fra le attività vietate nel settore della giustizia penale due specifici impieghi dei sistemi di intelligenza artificiale³³.

La prima attività vietata³⁴ riguarda l'utilizzo degli algoritmi di *risk assessment*³⁵, ad esclusione di quelli volti a fornire valutazioni basate su fatti oggettivi e verificabili, direttamente connessi a un'attività criminosa.

Si intende qui condividere il punto di vista dottrinale per cui la formula relativa ai «fatti oggettivi e verificabili» sia di scarsa efficacia e rischi di ricondurre l'utilizzo degli strumenti di

²⁵ Letteralmente: «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, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali».

²⁶ M. GIANGRECO, *Gli spazi applicativi dell'intelligenza artificiale nella giustizia penale: riflessioni alla luce dell'AI Act dell'UE e del Blueprint statunitense*, in *Cammino Diritto, Rivista di Informazione Giuridica*, n. 3/2025, p. 5.

²⁷ S. LORUSSO, *La sfida dell'intelligenza artificiale al processo penale nell'era digitale*, in *Sistema penale*, 2024, p. 1 ss.

²⁸ G. CANZIO, *AI Act e processo penale: sfide e opportunità*, in *Sistema penale*, 2024, p. 2.

²⁹ In *Ibidem*, p. 4.

³⁰ Regolamento (UE) 1689/2024, *cit.*, art. 13.

³¹ Regolamento (UE) 1689/2024, *cit.*, art. 14.

³² Regolamento (UE) 1689/2024, *cit.*, art. 27.

³³ I divieti di cui all'art. 5 del Regolamento hanno trovato applicazione a decorrere dal 2 febbraio 2025.

³⁴ Regolamento (UE) 1689/2024, art. 5, lett. d): «d'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di un sistema di IA per effettuare valutazioni del rischio relative a persone fisiche al fine di valutare o prevedere il rischio che una persona fisica commetta un reato, unicamente sulla base della profilazione di una persona fisica o della valutazione dei tratti e delle caratteristiche della personalità; tale divieto non si applica ai sistemi di IA utilizzati a sostegno della valutazione umana del coinvolgimento di una persona in un'attività criminosa, che si basa già su fatti oggettivi e verificabili direttamente connessi a un'attività criminosa».

³⁵ E. GUIDO, *Intelligenza artificiale e procedimento penale: ragionando di valutazione del rischio de libertate*, in *Archivio penale*, 2023, n. 1.

predictive policing vietati a valutazioni soggettive nonostante la preclusione dalla prima parte della disposizione in esame³⁶.

In questo senso, nella valutazione del rischio che un soggetto possa commettere un reato, l'algoritmo predittivo dovrebbe escludere ogni apprezzamento di dati soggettivi - fra cui i precedenti penali - altrimenti sfociando nel diritto penale d'autore³⁷, oltre a entrare in contrasto con il principio di presunzione di innocenza.

Inoltre, risulta immediatamente evidente che la norma sia indirizzata al divieto di utilizzo degli algoritmi di *risk assessment* nella valutazione di una singola persona, escludendo ogni tipo di riferimento al divieto di applicazione su di un gruppo di persone o su di un'intera comunità³⁸.

L'altra attività vietata riguarda l'utilizzo dei sistemi di identificazione biometrica remota «in tempo reale»³⁹ in luogo pubblico a fini di attività di contrasto⁴⁰, anche qui con specifiche eccezioni, a vario titolo riguardanti il contrasto a specifici crimini⁴¹.

Tale decisione è coerente con quanto già sostenuto dall'Agenzia europea per i diritti fondamentali (FRA), la quale caldeggiava un utilizzo «strictly determined» e limitato dell'utilizzo dello strumento di riconoscimento biometrico⁴².

Ulteriore limite è posto con riferimento all'obbligo di ottenere specifica autorizzazione da un'autorità giudiziaria indipendente dello Stato membro per poter svolgere attività di

³⁶ In M. GIANGRECO, *op. cit.*, p. 5; E. SACCHETTO, *Spunti per una riflessione sul rapporto tra biometria e processo penale*, in *Diritto penale contemporaneo*, n. 2/2019, p. 465 ss.

³⁷ In tal senso dispone il considerando n. 42, secondo cui le persone fisiche nell'Unione Europea «dovrebbero sempre essere giudicate in base al loro comportamento effettivo [e] non dovrebbero mai essere giudicate sulla base di un comportamento previsto dall'IA basato unicamente sulla profilazione, sui tratti della personalità o su caratteristiche quali la cittadinanza, il luogo di nascita, il luogo di residenza, il numero di figli, il livello di indebitamento o il tipo di automobile, senza che vi sia un ragionevole sospetto che la persona sia coinvolta in un'attività criminosa sulla base di fatti oggettivi verificabili e senza una valutazione umana al riguardo».

³⁸ Di questo avviso M. GIANGRECO, *op. cit.*, pag. 6, il quale ritiene che «questa possibile lettura ermeneutica permetterebbe di superare l'apparente contraddizione con l'Allegato III, che, nel legittimare i *risk assessment tools* «per valutare i tratti e le caratteristiche della personalità», sembra riferirsi alla possibile applicazione nei confronti di più soggetti, ai sensi della lett. d) dell'art. 5 *AI Act*. L'adozione di *risk assessment tools*, inoltre, potrebbe anche porsi in contrasto con un altro principio fondamentale dell'Unione Europea, che è quello di presunzione di innocenza».

³⁹ Sull'utilizzo di strumenti di identificazione biometrica, si rimanda a: E. SACCHETTO, *La prova biometrica*, in A. MARANDOLA, C. CONTI (a cura di), *La prova scientifica*, Milano, 2023, p. 243 ss.; G. MOBILIO, *Tecnologie di riconoscimento facciale. Rischi per i diritti fondamentali e sfide regolative*, Napoli, 2021; J. DELLA TORRE, *Norità dal Regno Unito: il riconoscimento facciale supera il vaglio della High Court of Justice (ma raccoglie le critiche del Garante Privacy D'Oltremare)*, in *Sistema penale*, 2020, p. 3.

⁴⁰ Per approfondimenti, si consiglia E. SACCHETTO, *Face to face: il complesso rapporto tra automated facial recognition technology e processo penale*, in *La legislazione penale*, 2020; B. PIETROCARLO, *Predictive policing: criticità e prospettive dei sistemi di identificazione dei potenziali criminali*, in *Sistema penale*, 2023.

⁴¹ Regolamento (UE) 1689/2024, *cit.*, art. 5, lett. h): «l'uso di sistemi di identificazione biometrica remota «in tempo reale» in spazi accessibili al pubblico a fini di attività di contrasto a meno che, e nella misura in cui, tale uso sia strettamente necessario per uno degli obiettivi seguenti: i) la ricerca mirata di specifiche vittime di sottrazione, tratta di esseri umani o sfruttamento sessuale di esseri umani, nonché la ricerca di persone scomparse; ii) la prevenzione di una minaccia specifica, sostanziale e imminente per la vita o l'incolumità fisica delle persone fisiche o di una minaccia reale e attuale o reale e prevedibile di un attacco terroristico; iii) la localizzazione o l'identificazione di una persona sospettata di aver commesso un reato, ai fini dello svolgimento di un'indagine penale, o dell'esercizio di un'azione penale o dell'esecuzione di una sanzione penale per i reati di cui all'allegato II, punibile nello Stato membro interessato con una pena o una misura di sicurezza privativa della libertà della durata massima di almeno quattro anni.»

⁴² European Union Agency for Fundamental Rights (FRA), *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, p. 25.

contrasto in luogo pubblico ma, anche in questo caso, la lettera della norma riferisce le eccezioni relative ai casi una situazione di urgenza debitamente giustificata⁴³.

Va in ogni caso sottolineato che la formulazione letterale delle eccezioni al generale divieto di utilizzo può risultare generica e, secondo la dottrina, tale difetto rischia di determinare disomogeneità applicativa nei diversi Stati membri, in particolare, circa la ricerca e l'individuazione degli autori di determinati reati⁴⁴.

Passando alle attività ritenute ad «alto rischio»⁴⁵ con riferimento al settore della giustizia penale, l'Allegato III⁴⁶ qualifica come tali i «sistemi usati per assistere un'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge a un insieme concreto di fatti».

È opportuno soffermarsi sulla portata di questa formula che, se letta da sola, non è scevra di potenziali fraintendimenti.

Il riferimento è ovviamente all'uso dei sistemi di IA per scopi ermeneutici che minano la genuinità della decisione⁴⁷ e limitano il principio fondamentale del libero convincimento del giudice⁴⁸, a causa dell'opacità che caratterizza i sistemi di giustizia predittiva⁴⁹.

Tale caratteristica, fra l'altro, a detta della Corte di giustizia, è legata al fatto che potrebbe «risultare impossibile comprendere la ragione per la quale un dato programma sia arrivato ad un riscontro positivo» e, dunque, «l'uso di siffatte tecnologie potrebbe privare gli interessati anche del loro diritto a un ricorso giurisdizionale effettivo sancito dall'articolo 47 della Carta [dei diritti fondamentali dell'Unione Europea]»⁵⁰.

Con riferimento a questo tipo di sistemi, la disciplina regolamentare prevede una serie di obblighi specifici, fra cui quello di predisporre e mantenere una documentazione tecnica dettagliata, la trasparenza per gli utenti – di cui si dirà –, la necessaria sottoposizione al controllo umano e l'adozione di misure *ad hoc* per i relativi rischi legati all'utilizzo del sistema.

Va detto che dal corpo della norma di cui all'articolo 9 del Regolamento si ricava quello che, come riportato in rubrica, è il «sistema di gestione dei rischi», consistente in un'articolata

⁴³ Regolamento (UE) 1689/2024, *cit.*, art. 5 par. 3: «Ai fini del paragrafo 1, primo comma, lettera h), e del paragrafo 2, ogni uso di un sistema di identificazione biometrica remota «in tempo reale» in spazi accessibili al pubblico a fini di attività di contrasto è subordinato a un'autorizzazione preventiva rilasciata da un'autorità giudiziaria o da un'autorità amministrativa indipendente, la cui decisione è vincolante, dello Stato membro in cui deve avvenire l'uso, rilasciata su richiesta motivata e in conformità delle regole dettagliate del diritto nazionale di cui al paragrafo 5. Tuttavia, in una situazione di urgenza debitamente giustificata, è possibile iniziare a usare il sistema senza autorizzazione a condizione che tale autorizzazione sia richiesta senza indebito ritardo, al più tardi entro 24 ore. Se tale autorizzazione è respinta, l'uso è interrotto con effetto immediato e tutti i dati nonché i risultati e gli *output* di tale uso sono immediatamente eliminati e cancellati.»

⁴⁴ In V. VASTA, *Diritto dell'Unione europea e intelligenza artificiale. Riflessi sul procedimento penale*, in AA. VV. (a cura di), *Intelligenza Artificiale. Diritto, giustizia, economia ed etica, Pubblicazioni del Dipartimento di Scienze Giuridiche "Cesare Beccaria"*, Bari, 2025, p. 127 ss., l'A ritiene che difficoltà applicative possano in particolare rinvenirsi nella definizione di un utilizzo relativo ad un nucleo di reati individuati sulla base del limite massimo edittale della pena, potendo questo mutare nelle varie legislazioni nazionali.

⁴⁵ La disciplina relativa ai sistemi ad «alto rischio» di cui all'art. 6, par. 1, dell'*AI Act* si applicano a decorrere dal 2 agosto 2027.

⁴⁶ Reg. (UE) 1689/2024, *cit.*, Allegato III, punto 8, lett. a).

⁴⁷ A. MACERATINI, *La giustizia predittiva: potenzialità e incognite*, in *MediaLAWs*, n. 2/2024, p. 239.

⁴⁸ In merito a tali riflessioni, si consiglia: R. E. KOSTORIS, *Intelligenza artificiale, strumenti predittivi e processo penale*, in *Cassazione penale*, n. 5/2024, p. 1642 ss.; G. UBERTIS, *Necessaria compatibilità dell'intelligenza artificiale con il giusto processo*, in *Archivio penale*, 2024; A. BALSAMO, *L'impatto dell'intelligenza artificiale nel settore della giustizia*, in *Sistema penale*, 2024; S. LORUSSO, *La sfida dell'intelligenza artificiale al processo penale nell'era digitale*, in *Sistema penale*, 2024.

⁴⁹ G. BARONE, *Giustizia predittiva e certezza del diritto*, Pisa, 2024.

⁵⁰ Corte di Giustizia UE (Grande Camera), sentenza del 21 giugno 2022, C-817/19, ECLI:EU:C:2022:491.

struttura giuridica composta da cautele, informazioni, garanzie, prove, sviluppi e progettazioni specifiche da adottare «al fine di eliminare o ridurre i rischi connessi all'uso del sistema di IA ad alto rischio».

A chiosa, lo stesso considerando n. 61 esclude dal novero dei sistemi di intelligenza artificiale ad «alto rischio», quelli «destinati ad attività amministrative puramente accessorie, che non incidono sull'effettiva amministrazione della giustizia nei singoli casi, quali l'anonimizzazione o la pseudonimizzazione di decisioni, documenti o dati giudiziari, la comunicazione tra il personale, i compiti amministrativi»⁵¹.

Come si vedrà di seguito, il sistema predisposto dall'AI Act è volto a promuovere un intervento di tipo "proattivo" più che "reattivo"⁵².

Tale tipo di approccio, infatti, è legato all'imposizione nei confronti dei soggetti coinvolti di una serie di obblighi e attività che danno vita ad un articolato sistema di responsabilità⁵³ che, esulando in ogni caso dalla stringente disciplina penale, permette di delimitare una zona bianca in cui considerare comunque lecito l'utilizzo di strumenti IA.

4. Il dovere di trasparenza: reale garanzia?

Fra gli obblighi più stringenti e potenzialmente più influenti nel settore della giustizia vi è l'obbligo di trasparenza, più volte inserito nel corpo del testo del Regolamento, quale componente necessaria per la conformità dei sistemi ad «alto rischio», da intendersi quale dovere di informare in modo chiaro, dettagliato e comprensibile gli utenti sul funzionamento dell'IA, specialmente in ambiti come la sorveglianza biometrica o la creazione di contenuti generati artificialmente⁵⁴.

Tale impegno è stato di recente ribadito anche nella Strategia europea in materia di giustizia elettronica, ove si legge che «spostare l'attenzione sulla prospettiva delle persone e rendere i sistemi giudiziari più accessibili, efficaci e trasparenti sarà [...] essenziale per rafforzare la fiducia tra le persone e le istituzioni pubbliche»⁵⁵.

In senso più pratico, deve essere assolutamente chiaro che l'utente sta interagendo con un sistema IA e lo stesso deve avere facile accesso alle informazioni relative al funzionamento del modello e ai criteri che guidano le decisioni automatizzate⁵⁶.

Dal considerando n. 27, invero, si ricava l'interpretazione del concetto di «trasparenza» nel senso che «i sistemi di IA sono sviluppati e utilizzati in modo da consentire un'adeguata tracciabilità e spiegabilità, rendendo gli esseri umani consapevoli del fatto di comunicare o

⁵¹ M. TORRE, *Il Regolamento europeo sull'intelligenza artificiale: profili processuali*, in *Processo penale e giustizia*, 2024, n. 6, p. 1534 ss.

⁵² J. DELLA TORRE, *Quale spazio per i tools di riconoscimento facciale nella giustizia penale?*, in G. DI PAOLO, L. PRESSACCO (a cura di), *Intelligenza artificiale e processo penale. Indagini, prove, giudizio*, Napoli, 2022, p. 20 ss.

⁵³ A tal proposito, per un quadro generale si veda M. E. FLORIO, *Il dibattito sulla responsabilità penale diretta delle I.A.: "molto rumore per nulla"?*, in *Sistema penale*, n. 2/2024.

⁵⁴ G. ZICCARDI, *op. cit.*, p. 27 ss.

⁵⁵ Strategia europea in materia di giustizia elettronica 2024-2028 (C/2025/437), in G.U.U.E., 16 gennaio 2025, C/1, Capo II-Principi, Par. A Principi sostanziali, lettera c) centralità delle persone.

⁵⁶ S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche?*, *Consolidati paradigmi normativi vs rischi e paure della giustizia digitale*, in *Cassazione penale*, 2019, pp. 1748-1765.

interagire con un sistema di IA e informando debitamente i *deployer*⁵⁷ delle capacità e dei limiti di tale sistema di IA e le persone interessate dei loro diritti».

Con riferimento alla «tracciabilità», si è già detto che per i sistemi ad «alto rischio» è fatto obbligo di predisporre documentazione tecnica che garantisca la conformità del sistema e che sia dettagliata in merito al funzionamento dell'algoritmo, sui dati utilizzati per l'addestramento e sui meccanismi di mitigazione del rischio.

Circa il profilo della «spiegabilità» del sistema pare utile soffermarsi, ben comprendendosi che l'uso dell'intelligenza artificiale nel settore della giustizia potrebbe avere un ruolo decisivo per i cittadini nella chiarificazione di concetti giuridici non di immediata comprensione.

In questo senso, la dottrina è giunta a riconoscere quattro caratteri chiave del concetto di «spiegabilità»: l'interpretabilità, la trasparenza, la giustificabilità e la riproducibilità⁵⁸.

Si ricava, pertanto, che per i sistemi di IA ad «alto rischio» è prevista la necessità di fornire informazioni appropriate agli utilizzatori, mediante istruzioni per l'uso che coprano le caratteristiche, le prestazioni, i limiti del sistema, i rischi possibili e le misure di controllo umano pertinenti, anche attraverso esempi esplicativi⁵⁹.

Non a caso, in ossequio ai dettami di cui all'articolo 96 del Regolamento, rubricato «Orientamenti della Commissione sull'attuazione del regolamento», nel luglio del 2025 la Commissione europea ha pubblicato il Codice di buone pratiche sull'IA per finalità generali, cui si affiancano orientamenti sui concetti chiave relativi ai modelli di intelligenza artificiale per finalità generali (GPAI).

Limitandoci qui a riferire del carattere volontario di tale Codice, la Commissione ha allegato, nel capo sulla trasparenza, un «modulo di documentazione tipo» che potrà consentire ai fornitori di modelli GPAI di documentare facilmente le informazioni necessarie in conformità al Regolamento sull'IA⁶⁰.

In effetti, dal considerando 72 non era chiaro in che modo sia in concreto possibile descrivere e spiegare parametri tecnici nel linguaggio naturale, quale livello di chiarezza e precisione nella spiegazione sia necessario e come sia possibile valutarne il raggiungimento⁶¹.

⁵⁷ Tale termine, non tradotto nella versione italiana del Regolamento, intende, secondo l'art. 3, par. 4: «una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che utilizza un sistema di IA sotto la propria autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale».

⁵⁸ In *ibidem*, p. 22, l'A. precisa i contenuti di ciascun carattere: «i) la interpretabilità (riguarda la possibilità di comprendere le relazioni tra *input* e *output* di un modello di IA, e alcuni modelli, come le reti neurali profonde, sono difficili da interpretare, mentre altri, come gli alberi decisionali, sono più trasparenti); ii) la trasparenza (si riferisce alla capacità di esaminare il funzionamento interno di un sistema di IA, come i pesi e i parametri di un modello di apprendimento automatico, e anche in questo caso alcuni algoritmi sono considerati «scatole nere» perché il loro processo decisionale è opaco e complesso); iii) la giustificabilità (implica la possibilità di fornire motivazioni coerenti e logiche che spieghino una decisione presa dal sistema, in modo comprensibile per gli esseri umani), e iv) la riproducibilità (significa che un sistema di IA dovrebbe produrre risultati simili quando vengono forniti dati e condizioni simili, permettendo di verificare e validare il suo funzionamento)».

⁵⁹ M. R. ALLEGRI, *L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio*, in *MediaLAWs*, n. 3/2024, pp. 19-27.

⁶⁰ Commissione europea, *Transparency Chapter*, in *Code of Practice for General-Purpose AI Models*, del 10 luglio 2025.

⁶¹ Dal testo si ricava che: «i fornitori dovrebbero garantire che tutta la documentazione, comprese le istruzioni per l'uso, contenga informazioni significative, complete, accessibili e comprensibili, tenendo conto delle esigenze e delle conoscenze prevedibili dei *deployer* destinatari. Le istruzioni per l'uso dovrebbero essere messe a disposizione in una lingua che possa essere compresa facilmente dai *deployer* destinatari, secondo quanto stabilito dallo Stato membro interessato».

L'articolo 11 lascia analoghe zone di incertezza applicativa nel disciplinare che i sistemi di IA ad alto rischio devono essere corredati dalla documentazione tecnica prevista – in maniera molto specifica – dall'Allegato IV del Regolamento, da fornire alle autorità nazionali competenti e agli organismi notificati.

Tali perplessità sono emerse in particolare sulla possibilità di tradurre concretamente le regole tecniche dal linguaggio computazionale al linguaggio naturale, per renderle comprensibili a persone non necessariamente esperte del linguaggio matematico-informatico⁶².

Come già accennato, infatti, l'Unione europea ha interesse a perseguire parallelamente sia lo sviluppo dello strumento di IA nel rispetto dei principi fondamentali, sia l'accessibilità da parte dei cittadini in termini di trasparenza e comprensibilità⁶³.

Tale volontà legislativa si è tradotta, nell'ambito della giustizia, nella scelta di sfruttare lo strumento informatico nella duplice ottica tutoria dell'autorità giudiziaria e degli addetti all'amministrazione della giustizia, e informativa e conoscitiva per i cittadini⁶⁴.

In questo senso, il dovere di trasparenza imposto dovrebbe garantire l'utilizzo di un linguaggio accessibile e l'estrapolazione di risultati giurisprudenziali, normativi e processuali corretti per i cittadini, ma soprattutto per giudici, avvocati e professionisti del diritto.

Detto che i benefici derivanti da un tale uso sarebbero di assoluto pregio sotto tanti profili, non ne vanno sottaciuti i rischi ancora attuali⁶⁵.

In questo senso, gli strumenti di "riassumi con IA" o "semplifica con IA" che molte banche dati giuridiche online hanno già adottato permette un'estrazione della conoscenza certamente più accessibile per il cittadino, ma potenzialmente meno efficace e, in alcuni casi, incompleta per gli esperti.

In effetti, la lunghezza dei testi legislativi, delle decisioni giudiziarie e dei procedimenti amministrativi è già notevolmente complessa per gli esperti del settore ma l'uso dello strumento IA può certamente contribuire a creare un ecosistema di fiducia e promuovere un avvicinamento fra tutte le parti.

D'altronde la legge entra ogni giorno nella vita di ciascuno di noi e, anche per evitare spiacevoli inconvenienti⁶⁶, l'algoritmo che propone una soluzione tecnica dovrebbe comunque, in un'ottica di trasparenza, suggerire il confronto con il testo giuridico integrale o la consultazione di un esperto⁶⁷.

In tal senso, potrebbe essere utile lo sviluppo nel settore giuridico della c.d. "intelligenza artificiale spiegabile" (XAI – *Explainable Artificial Intelligence*), strumento che ha

⁶² M. R. ALLEGRI, *op. cit.*, p. 24.

⁶³ In questo senso, si veda: M. FRANCAVIGLIA, *IA e funzioni giurisdizionali: alcune questioni preliminari alla luce del quadro costituzionale*, in *Rivista italiana di informatica e diritto*, n. 1/2025, pp. 1-16.

⁶⁴ S. QUATTROCOLO, *Intelligenza Artificiale e processo penale: le novità dell'AI Act*, in *Diritto di Difesa, la Rivista dell'Unione delle camere penali italiane*, pp. 1-13.

⁶⁵ G. UBERTIS, *Processo penale telematico, intelligenza artificiale e Costituzione*, in *Cassazione penale*, 2024, n. 2.

⁶⁶ Il riferimento è alla vicenda avvenuta presso il Tribunale di Firenze, in cui un Avvocato inseriva all'interno di un proprio atto una serie di riferimenti giurisprudenziali errati. Tale episodio rientra nel più ampio fenomeno delle c.d. "allucinazioni di intelligenza artificiale".

Per una lettura completa della vicenda, si veda: <https://www.altalex.com/documents/2025/04/01/sentenza-viene-inventata-a-i-responsabilita-ex-art-96-c-p-c>.

⁶⁷ G. VACIAGO (a cura di), *Intelligenza artificiale generativa e professione forense*, Milano, 2024.

l'obiettivo di migliorare l'interpretabilità dei modelli AI e che punta alla fiducia e alla sicurezza degli utenti⁶⁸.

Con tale locuzione si intende un insieme di tecniche pensate per rendere interpretabili gli algoritmi, offrendo una spiegazione chiara e standardizzata del percorso logico che conduce a un determinato *output*, chiarendo quali variabili hanno influenzato la decisione e con quale peso⁶⁹.

Condividendosi l'indirizzo dottrinale secondo cui in settori sensibili deve pretendersi un grado di comprensibilità e un livello di spiegazione molto alti, pare opportuno specificare che l'implementazione del modello di IA spiegabile nel settore della giustizia potrebbe avere riflessi importanti in nuove declinazioni del diritto costituzionale alla difesa⁷⁰.

In tal senso, pare ci sia ancora da fare per eliminare incertezza teorica e pratica di tali disposizioni.

La sfida è aperta e spetta agli Stati e alle istituzioni europee mettere in pratica i principi etici contenuti nel Regolamento⁷¹, i quali, nell'attuale contesto tecnologico e di mercato, rischiano di rimanere inapplicati o, ancor peggio, fumosi.

5. Conclusioni

Il quadro normativo europeo in materia di *AI Act* ha certamente segnato un importantissimo passo in avanti nel tentativo di incardinare lo sviluppo tecnologico all'interno di un perimetro giuridico chiaro e strutturato.

Allo stesso tempo però, come messo in evidenza, la disciplina in oggetto sconta ancora un margine di indeterminatezza che mette a rischio la tutela dei diritti fondamentali garantiti dal Regolamento.

In questo senso, è chiaro a tutti, istituzioni e *stakeholders*, che il percorso di introduzione dei sistemi di Intelligenza Artificiale nel settore della giustizia non può essere immediato e passivo.

Al contrario, dovrà essere accompagnato da un processo di adattamento graduale di carattere culturale *in primis*, ma anche formativo e giuridico, per garantire il principio base dell'antropocentrismo.

⁶⁸ Per un approfondimento sulle varie tecniche messe a punto da coloro i quali si occupano di *Explainable LA* si suggerisce: A. LOREGGIA, G. SARTOR, *L'Intelligenza Artificiale nella moderazione del digitale*, in *Sistemi Intelligenti*, pp. 53-57; P. BUONO, G. DESOLA, R. LANZILOTTI, *Interazione con sistemi intelligenti*, in V. V. CUOCCI, F. P. LOPS, C. MOTTI (a cura di), *La responsabilità civile nell'era digitale: atti della Summer school 2021*, Bari, 2022, p. 115 ss.

⁶⁹ G. CONTISSA, G. LASAGNI, G. SARTOR, *Quando a decidere in materia penale sono (anche) algoritmi e LA: alla ricerca di un rimedio effettivo*, in *Dirittodiinternet.it*, n. 4/2019.

⁷⁰ C. CASONATO, *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, in *BioLaw Journal – Rivista di BioDiritto, Special Issue*, n. 2/2019, pp. 722-724. L'Autore inquadra tre importanti diritti di rango costituzionale, in particolare: «a) [...] ottenere una spiegazione dei passaggi attraverso i quali la macchina ha generato il proprio risultato; b) [...] essere resi consapevoli della natura, umana o artificiale, del proprio interlocutore; c) [...] essere destinatari di decisioni che siano il risultato di un processo in cui sia presente una significativa componente umana».

⁷¹ Strategia europea in materia di giustizia elettronica 2024-2028, in *GUUE*, C/2025/437, del 16 gennaio 2025.

Pare opportuno riferire che nei primi mesi del 2025, in ossequio all'articolo 96 del Regolamento, la Commissione ha pubblicato la bozza delle Linee Guida afferenti alle pratiche vietate di cui all'articolo 5⁷².

Tale disposizione impone alla Commissione europea di procedere nelle medesime forme anche sulla disciplina relativa all'utilizzo dei sistemi identificati ad «alto rischio» dall'art. 6 del Regolamento.

Interessante dunque sarà comprendere come la Commissione suggerirà di procedere nell'ambito dell'amministrazione della giustizia per garantire i diritti fondamentali dinanzi al pericolo della disumanizzazione.

Infatti, il rischio potrebbe essere che l'introduzione dei modelli di IA a supporto del giudice, in assenza di linee guida chiare e adattabili a tutti gli Stati membri, determini un *vulnus* di tutela negli ordinamenti giuridici europei, con importanti conseguenze sull'accesso alla giustizia dei cittadini e sul rispetto dei principi fondamentali delle persone imputate⁷³.

In questo senso, si è voluto sottolineare il ruolo del principio di trasparenza, spesso evocato all'interno del Regolamento, in quanto ritenuto un ottimo punto di partenza per la chiarezza comunicativa, l'accessibilità tecnica e la garanzia del rispetto dell'eticità su cui è imperniata la disciplina.

Tuttavia, il rischio, come sottolineato in apertura e ripreso dalla provocazione "dallesca", è quello di trovarsi di fronte a un'infrastruttura giuridica tanto potente quanto opaca, una "anidride solforosa" normativa che illude con la promessa della modernità ma che, in mancanza di un'effettiva applicazione dei suoi principi, potrebbe accentuare le distorsioni anziché ridurle.

⁷² Commissione europea, Annex to the Communication to the Commission, Approval of the content of the draft Communication from the Commission - Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (*AI Act*), del 4 febbraio 2025, C(2025) 884 final.

⁷³ L. PALMIERI, *Il deposito telematico tra esigenze efficientiste e garanzie difensive*, in *Diritto penale e processo*, n. 4/2024.



ARIANNA BAVASTRELLI*

INTELLIGENZA ARTIFICIALE E GIUSTIZIA: ALCUNE PROSPETTIVE IN AMBITO UE

SOMMARIO: 1. Introduzione. – 2. Lo stato dell'arte della digitalizzazione nell'Unione europea. – 3. I più recenti sviluppi normativi: *AI Act*. – 4. Intelligenza artificiale e giustizia. – 5. Il controllo umano dei processi operativi nel Regolamento (UE) 2024/1689. – 6. Conclusioni.

1. Introduzione

Le radicali trasformazioni introdotte dalla tecnologia digitale esercitano, oggi, un impatto molto profondo nei diversi settori del diritto. I rischi correlati ad una sua applicazione, in ambito giuridico, non possono essere sottovalutati, poiché le problematiche che ne derivano sono molteplici e di notevole complessità.

Il presente contributo tenterà di delineare solo alcuni dei più delicati profili legati all'utilizzo dei sistemi di intelligenza artificiale all'interno dei procedimenti giudiziari.

È ormai noto come i correnti progressi tecnologici riescano ad intersecarsi con qualsivoglia attività tipicamente umana per cui, soprattutto in settori come quello della giustizia, si rende necessario comprendere sia le modalità ed i limiti con cui la digitalizzazione potrà incidere sul processo, sia come quest'ultimo si adeguerà agli sviluppi tecnologici ormai ampiamente diffusi¹.

L'indagine che si intende proporre muoverà dalle iniziative promosse dall'Unione europea nell'ambito della transizione digitale attualmente in corso il cui obiettivo, da oltre un ventennio, si sostanzia nella realizzazione di un mercato unico digitale che consenta la libera circolazione di dati preservando, al contempo, il rispetto dei diritti fondamentali degli individui.

Infatti, un impiego incontrollato dei sistemi di intelligenza artificiale, per esempio, potrebbe minare il rispetto delle garanzie multilivello, nonostante l'importanza che questi

* Dottoranda di ricerca in Studi europei, Dottorato di Interesse Nazionale (DIN) dell'Università di Genova, *Curriculum Governance multilivello e diritti fondamentali*, sede di afferenza Università degli Studi di Messina, Dipartimento di Scienze Politiche e Giuridiche.

¹ A. PAJNO, *L'uso dell'intelligenza artificiale nel processo tra problemi nuovi e questioni antiche*, in *BioLaw Journal – Rivista di BioDiritto*, n. 1/2022.

nuovi strumenti continuano ad acquisire per una maggiore crescita economica e una migliore qualità della vita².

La crescente attenzione che ruota attorno alle applicazioni di intelligenza artificiale, soprattutto per ciò che attiene ad un loro utilizzo nei procedimenti giudiziari, condurrà l'analisi sulla più recente normativa adottata, a livello europeo, in materia di IA: il Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, meglio conosciuto come *AI Act*.

È opportuno precisare sin da subito che l'atto normativo, in vigore dal primo agosto 2024, subirà un'applicazione graduale. Precisamente, la parte relativa ai sistemi vietati è entrata in vigore il 2 febbraio 2025. Diversamente, altre sezioni del Regolamento diverranno applicabili a partire dal 2 agosto 2025, dal 2 agosto 2026 e dal 2 agosto 2027³.

Con l'obiettivo di dotare l'Unione di una normativa più uniforme in materia, le istituzioni europee hanno elaborato una disciplina volta a superare le normative nazionali divergenti che potrebbero «determinare una frammentazione del mercato interno e diminuire la certezza del diritto per gli operatori che sviluppano o utilizzano sistemi di IA»⁴.

Se si tiene conto che il Regolamento (UE) 2024/1689, come si avrà modo di riprendere meglio in seguito, inserisce il settore della giustizia fra i sistemi dallo stesso definiti ad «alto rischio», ben si comprende la necessità, avvertita dalle istituzioni sovranazionali, di dare vita ad un quadro normativo non solo innovativo ma, allo stesso tempo, in grado di stabilire uno *standard* elevato per la tutela dei diritti fondamentali.

In ambito giudiziario, profili legati, per esempio, alla giustizia predittiva rappresentano una rivoluzione capace di incidere su piani diversi, persino mettendo in discussione l'apporto umano nell'attività di *ius dicere*⁵.

Inoltre, il Regolamento chiarisce il ruolo demandato all'intervento umano nei diversi utilizzi di strumenti di IA, con particolare riferimento agli impieghi di sistemi ad «alto rischio» idonei ad incidere sulla sicurezza e sui diritti fondamentali.

Introdotta sulla scorta di un approccio *human oriented*, la sorveglianza umana è calibrata sull'intensità e sulla tipologia dei rischi prodotti dall'intelligenza artificiale e dalle tecnologie ad essa correlate.

Infine, anche l'aspetto legato alla velocità evolutiva che si registra in tema di digitalizzazione e, ancor di più, di intelligenza artificiale condurrà l'Unione europea a monitorare gli effetti derivanti dall'applicazione della normativa sopra richiamata, una volta che questa diverrà integralmente applicabile.

Centrale, sul punto, l'apporto della funzione interpretativa che la Corte di giustizia è chiamata a svolgere quale garante dei valori sanciti a livello sovranazionale per la tutela dei diritti fondamentali degli individui.

² F. DONATI, *Diritti fondamentali e algoritmi nella proposta di regolamento sull'intelligenza artificiale*, in *Il Diritto dell'Unione Europea*, n. 3-4/2021, p. 453 ss.

³ Sul punto si rimanda a quanto previsto dall'articolo 113 del Regolamento per una maggiore completezza.

⁴ Dal considerando n. 3 del Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 *che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (Regolamento sull'intelligenza artificiale)*, in GUUE Serie L del 12/7/2024.

⁵ S. ARDUINI, *La "scatola nera" della decisione giudiziaria: tra giudizio umano e giudizio algoritmico*, in *BioLaw Journal – Rivista di BioDiritto*, n. 2/2021.

2. Lo stato dell'arte della digitalizzazione nell'Unione europea

L'Unione europea, da oltre un decennio, promuove fra i suoi principali obiettivi il più ampio sviluppo di tecnologie digitali finalizzato all'implementazione dei settori di interesse pubblico, così da incidere positivamente sia sul piano europeo, migliorando la cooperazione, sia all'interno degli ordinamenti statali, rafforzandoli dal punto di vista digitale.

A partire dai primi anni 2000, l'Unione ha posto una particolare attenzione sul tema presentando la prima agenda digitale europea (2010-2020)⁶ e, successivamente, il programma "Europa digitale" per il periodo 2021/2027⁷; questi primi interventi hanno dato impulso al processo di digitalizzazione definendo un quadro normativo in grado di rafforzare le capacità digitali dell'UE, facilitandone la diffusione, e promuovere la trasformazione digitale della società dell'Unione.

Questa ambizione assume oggi i caratteri di vera e propria esigenza soprattutto alla luce della crisi legata alla pandemia da Covid 19⁸.

Dal 2020, infatti, si è assistito ad un incisivo mutamento della funzione delle tecnologie digitali, il cui impatto ha influito drasticamente sulla società moderna di tutto il mondo, permettendole di superare limiti fisici e geografici.

Negli ultimi anni, infatti, la Commissione europea ha evidenziato che tra le finalità di maggiore rilievo, cui ambire entro il 2030, vi è sicuramente quella di raggiungere una sovranità digitale nei vari settori di competenza dell'Unione affinché, in piena armonia con gli Stati membri, si possa pervenire ad una normativa uniforme.

Invero, nell'ottica di un futuro digitale al servizio dei cittadini e delle pubbliche amministrazioni dell'Unione, il presente lavoro intende concentrare l'attenzione sul settore della giustizia penale e sulle modalità in cui questa potrà essere amministrata, alla luce delle recenti implicazioni derivanti dall'utilizzo dell'intelligenza artificiale⁹.

L'emergenza sanitaria richiamata ha portato alla luce le sfide legate al funzionamento efficace dei sistemi giudiziari in circostanze eccezionali nell'ottica di un conseguente rafforzamento di questi, sollecitando tanto il perseguimento di obiettivi in parte già programmati quanto richiedendo all'Unione europea degli interventi specifici e concreti.

L'articolo 67 del Trattato sul funzionamento dell'Unione europea sancisce che l'accesso alla giustizia e la cooperazione transfrontaliera devono essere garantiti all'interno dello spazio di libertà, sicurezza e giustizia ed è pacifico sostenere che, nell'era digitale, questi principi necessitano di essere salvaguardati anche alla luce del costante cambiamento e progresso digitale.

Tra le sfide attuali dell'ordinamento dell'Unione vi è sicuramente quella di progettare ed attuare un modello di giustizia sempre più digitale.

⁶ Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, *Un'agenda digitale europea*, del 19.5.2010 COM (2010)245 definitivo.

⁷ Regolamento (UE) 2021/694 del Parlamento europeo e del Consiglio del 29 aprile 2021 che istituisce il programma Europa digitale e abroga la decisione (UE) 2015/2240.

⁸ Sul ruolo della digitalizzazione nel contesto pandemico si richiama la Risoluzione del Parlamento europeo del 17 aprile 2020 sull'azione coordinata dell'UE per lottare contro la pandemia di COVID-19 e le sue conseguenze (2020/2616(RSP)), (2021/C 316/01), ed anche il documento del Consiglio europeo *Una tabella di marcia per la ripresa – Verso un'Europa più resiliente, sostenibile ed equa* approvato il 23 aprile 2020.

⁹ *Artificial Intelligence and Law Enforcement - Impact on Fundamental Rights Study* commissionato dal Dipartimento tematico per i diritti dei cittadini e gli affari costituzionali del Parlamento europeo su richiesta della Commissione LIBE del luglio 2020, <https://www.europarl.europa.eu/portal/>.

Allo stesso tempo, è necessario che tale esigenza si realizzi nel pieno rispetto dei diritti fondamentali poiché, alla luce dell'articolo 47 della Carta dei diritti fondamentali dell'Unione europea, diritti come il ricorso effettivo a un giudice imparziale o la protezione dei dati personali devono poter essere esercitati da tutte le persone, ivi compresi minori, anziani, soggetti affetti da disabilità e vittime di reato, in condizioni di assoluta parità ed ognuno in conformità alle proprie specifiche esigenze.

L'ammodernamento digitale dei sistemi giudiziari, sino ad ora, è dipeso quasi esclusivamente dalla volontà dei singoli Stati membri che, nell'assenza di disciplina comunitaria, avevano avviato in modo indipendente diversi progetti come, ad esempio, l'introduzione di piattaforme attraverso cui reperire dati normativi e orientamenti giurisprudenziali¹⁰.

3. I più recenti sviluppi normativi: l'AI Act

Alla luce del quadro sin qui delineato, si intende porre particolare attenzione sulle recenti difficoltà nell'approccio alle nuove potenziali problematiche relative all'intelligenza artificiale e al suo utilizzo all'interno dei procedimenti giudiziari.

La regolamentazione dell'intelligenza artificiale non è una priorità emersa di recente tra gli obiettivi dell'Unione europea.

A rendere il dibattito sul tema sempre più acceso è, senza dubbio, la velocità esponenziale con cui tali sistemi si evolvono e l'incremento della produzione normativa in materia.

Nel febbraio 2020, la Commissione europea ha adottato il Libro bianco sull'intelligenza artificiale¹¹ con l'impegno di favorire i progressi scientifici preservando il primato tecnologico dell'UE, alla luce della forte dipendenza da tecnologie non europee che non consentono agli Stati membri di convergere in materia di digitalizzazione, e di affrontare i rischi associati all'utilizzo di questa nuova tecnologia.

L'intento dell'Unione di rafforzare il proprio ruolo guida all'interno del panorama mondiale, in tema di intelligenza artificiale, ha trovato un primo riscontro con l'adozione del Regolamento (UE) 2024/1689¹² (*AI Act*), quale primo strumento giuridico esclusivamente incentrato sulla disciplina di questa tecnologia, il cui fine è stabilire un quadro giuridico armonizzato per l'uso dell'IA all'interno dell'Unione europea.

La linea normativa tracciata dal Regolamento non consiste nella previsione di regole uniformi per tutti i sistemi di IA, ma segue un approccio basato sul rischio, declinato in quattro differenti livelli, delineando di conseguenza diversi assetti giuridici.

Ad ognuno di questi sistemi corrispondono diversi obblighi correlati a rischi per gli interessi pubblici di riferimento e per i valori e i diritti fondamentali che l'ordinamento giuridico europeo tutela.

¹⁰ S. SIGNORIN, *Verso l'applicazione trasversale del principio "digital by default" nel settore della cooperazione giudiziaria dell'Unione europea* in *Eurojus*, n. 1/2025.

¹¹ Commissione europea, *Libro bianco sull'intelligenza artificiale – Un approccio europeo all'eccellenza e alla fiducia*, del 19.2.2020, COM (2020) 65 final.

¹² Secondo O. POLLICINO, in *Regolazione e innovazione tecnologica nell'ordinamento della rete*, in *Rivista AIC*, 2025, p. 127, può essere definito quale «primo esperimento globale».

Si potrebbe, dunque, affermare che l'obiettivo dello strumento adottato sia quello di sostenere lo sviluppo, l'estensione e la *governance* responsabile dei sistemi di IA nei territori dell'Unione stabilendo regole pertinenti, proteggendo i diritti fondamentali e favorendo altresì l'innovazione¹³.

È opportuno, inoltre, evidenziare che l'*AI Act*, sin dalla sua fase embrionale, mira a riflettere al meglio il binomio tra regolamentazione e innovazione digitale.

Già la Commissione aveva messo in luce come fosse indispensabile puntare alla creazione di un mercato unico per l'intelligenza artificiale in grado di superare la frammentazione e di offrire agli operatori, dei diversi settori, regole chiare e precise¹⁴.

Allo stesso tempo, il Regolamento si propone di disciplinare l'accesso esclusivo al mercato unico dell'Unione europea, ai sistemi di IA sicuri, affidabili ma soprattutto conformi ai valori europei, con specifico riferimento a quelli che promuovono la tutela dei diritti fondamentali, vale a dire quei sistemi in cui i diritti fondamentali rappresentano il fulcro attorno al quale la tecnologia può correttamente espandersi, in virtù di un'ormai consolidata visione antropocentrica.

Per tale ragione, la normativa intende rimediare ad un'ulteriore ed eventuale frammentazione nella tutela stessa dei diritti, affinché gli Stati membri non impongano un differente *standard* di tutela dei diritti fondamentali in materia di IA nel quadro europeo, laddove, viceversa, potrebbero coesistere da un lato contesti di assoluto divieto di taluni strumenti, dall'altro scenari in cui vengono immessi sul mercato sistemi potenzialmente dannosi per la vita o la dignità degli individui¹⁵.

Nel rispetto delle competenze conferite dai Trattati, l'Unione europea ha adottato il Regolamento (UE) n. 1689/2024 attribuendo all'atto normativo portata orizzontale, destinato ad applicarsi ai diversi ambiti di IA a prescindere dal settore di riferimento.

Dal momento che la "sola" tutela dei diritti fondamentali, di per sé, non potrebbe costituire una base giuridica, la tutela *de quo* viene supportata dalla disciplina offerta dall'articolo 114 del Trattato sul funzionamento dell'Unione europea, quale disposizione che regola l'armonizzazione delle norme relative al mercato interno.

La versione definitiva del Regolamento ha subito delle modifiche rilevanti, nel corso dell'*iter* che ha condotto alla sua approvazione, dando vita ad un risultato nuovo sul piano normativo che favorisce lo stretto collegamento tra la definizione degli *standard* tecnici e la previsione di diritti e obblighi in capo agli sviluppatori dei sistemi di IA.

Ed invero, la dinamicità e la rapidità con cui questi sistemi si evolvono hanno reso necessaria l'adozione di una disciplina più flessibile caratterizzata da disposizioni generiche e meno rigide, al fine di eludere il rischio di una rapida obsolescenza del Regolamento che risulterebbe, viceversa, vincolato da regole fin troppo circoscritte.

Ne consegue un ampliamento della funzione interpretativa, demandata, in larga misura, all'intervento della giurisprudenza¹⁶.

¹³ M. PERLOT, *Credit scoring: il bilanciamento dei diritti fondamentali nell'AI Act*, in *Quaderni AISDUE*, anticipazione fascicolo *Intelligenza artificiale e nuove vulnerabilità*.

¹⁴ Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, *L'intelligenza artificiale per l'Europa*, del 25 aprile 2018 COM (2018) 237final.

¹⁵ C. SCHEPISI, *Brevi note sulla "dimensione europea" della regolamentazione dell'intelligenza artificiale: principi, obiettivi e requisiti*, in V. FALCE (a cura di), *Strategia dei dati e intelligenza artificiale. Verso un nuovo ordine giuridico del mercato*, Torino, 2023, p. 53 ss.

¹⁶ E. CIRONE, *L'AI Act e l'obiettivo (mancato?) di promuovere uno standard globale per la tutela dei diritti fondamentali*, in *Quaderni AISDUE* n. 2/2024.

4. Intelligenza artificiale e giustizia

Concentrando l'analisi sull'impatto che l'intelligenza artificiale può avere nel campo della giustizia, è necessario partire dal dato normativo di cui al punto 8 dell'Allegato III dell'*AI Act*, vale a dire la scelta di inserire tra i sistemi di intelligenza artificiale ad «alto rischio» quelli «destinati a essere usati da un'autorità giudiziaria o per suo conto per assistere un'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti»¹⁷.

Le motivazioni a sostegno di tale scelta emergono già dal Considerando n. 61 del Regolamento, laddove l'Unione manifesta la volontà di contenere il pericolo di “potenziali distorsioni, errori e opacità” tenuto conto che, più in generale, l'*AI Act* classifica come sistemi ad alto rischio quei particolari sistemi di intelligenza artificiale destinati all'amministrazione della giustizia e ai processi democratici¹⁸.

Quando si fa riferimento all'ingresso della IA nel settore del processo penale è doveroso muovere da una serie di rilevanti ed evidenti considerazioni: si tratta di un settore dell'ordinamento particolarmente sensibile quanto suscettibile di pericolose interferenze, dal momento che si giudica della libertà *ergo* dei diritti e delle garanzie riconosciuti agli individui e ad essa strettamente correlate.

Per tale ragione, ogni misura posta a garanzia delle possibili applicazioni dell'IA in qualsivoglia sfera del diritto, in questo specifico settore necessiterebbe di attenzioni ulteriori.

Volendo essere più chiari, nel bilanciamento di interessi in gioco, qualora le implementazioni di IA dovessero potenzialmente provocare compressioni dei diritti fondamentali, nessun vantaggio derivante dall'introduzione di tali strumenti potrebbe sufficientemente compensarli.

Ad ogni modo, sono diversi gli strumenti di IA di cui il processo penale potrebbe servirsi dal momento che il progresso tecnologico rende la gamma sempre più ampia e potrebbe continuare a farlo in futuro. Da dispositivi per la ricerca e l'accesso a dati e informazioni a strumenti che permetterebbero al giudice di farsi assistere nella ricostruzione logico deduttiva che lo conduce a sentenza.

Potrebbe pacificamente prevedersi che un corretto sviluppo di sistemi di intelligenza artificiale vengano finalizzati a rendere l'accesso dei cittadini alla giustizia più agevole, specialmente in tutti quei casi in cui si permetterebbe di abbattere le barriere culturali e linguistiche che ostano alla partecipazione, effettiva, di tutti gli attori del procedimento.

A titolo meramente esemplificativo, si osserva come le nuove tecnologie potrebbero consentire a chiunque di presentare correttamente *online*, in maniera guidata dal portale e dalle scelte che il singolo può effettuare sulla base delle proprie esigenze, atti del procedimento mediante la predisposizione di siti *web* da parte degli organi giurisdizionali. Tali circostanze, se si tiene conto della frequenza con cui si rende necessario prendere parte ad un

¹⁷ Punto 8 a Allegato III Sistemi di IA ad alto rischio di cui all'articolo 6, paragrafo 2 del Regolamento (UE) 2024/1689.

¹⁸ A. BALSAMO, *L'impatto dell'intelligenza artificiale nel settore della giustizia*, testo dell'audizione del Sostituto Procuratore generale presso la Corte di Cassazione Dott. Antonio Balsamo nell'ambito dell'indagine conoscitiva sull'impatto dell'intelligenza artificiale nel settore della giustizia, avvenuta il 12 marzo 2024 dinanzi alla Commissione Giustizia del Senato della Repubblica, <https://www.sistemapenale.it>.

procedimento transfrontaliero che si celebra in un Paese diverso da quello di residenza del soggetto, avrebbero quale fine ultimo quello di conseguire un risultato significativo in termini di efficienza.

Allo stesso tempo non può non tenersi conto della complessità delle implementazioni di tipo “predittivo” e degli strumenti che combinano questi ultimi con altri sistemi, incluse le applicazioni di IA di tipo “generativo”¹⁹.

In un’ottica più ampia, si potrebbe enucleare la tecnologia che viene in rilievo in ambito giudiziario in tre macro gruppi: la tecnologia di supporto, in grado di assistere nell’informare, supportare e consigliare tutti coloro che si ritrovano coinvolti nel sistema giudiziario; la tecnologia sostitutiva, che potrebbe svolgere funzioni e attività in precedenza svolte da esseri umani; infine la tecnologia dirompente può tener conto di una parte del processo decisionale umano²⁰.

Senza dubbio, tra i punti cruciali in tema di rischi legati all’impiego dell’IA nell’amministrazione della giustizia vi è il carattere predittivo che emerge sia con riferimento alla fase investigativa sia alla fase di valutazione probatoria e decisoria del giudice, tanto da rendersi necessaria l’adozione di estreme cautele.

Per quanto attiene alla seconda delle fasi richiamate, sarebbe opportuno non confondere la certezza del diritto con la prevedibilità delle decisioni emesse a seguito di un procedimento o con l’esigenza di rendere celere gli esiti di una controversia. Sovrapporre questi due aspetti potrebbe portare al sacrificio di una parte fondamentale dell’attività giurisdizionale in senso stretto, coincidente con il momento decisionale del caso concreto.

Ed il giudice, a differenza del legislatore che emana una disciplina generale ed astratta, è chiamato ad applicare principi di portata generale al caso di specie²¹.

Tuttavia, gli strumenti predittivi, in questo ambito, possono rilevare con riferimento a momenti ulteriori e fungere, talora, anche da indicatore del rischio soprattutto quando ci si riferisce ad attività di prevenzione criminale. In questo specifico contesto, che per definizione si pone non soltanto a monte dell’instaurarsi di un procedimento ma ancor prima della potenziale commissione di un determinato reato, gli strumenti predittivi consentirebbero di effettuare controlli preventivi su certi territori e nei confronti di categorie di soggetti ben specifiche, in maniera tale da svolgere opere di prevenzione del crimine che tenterebbero di soddisfare i requisiti della determinatezza, affidabilità ed efficienza²².

Ne deriva che le problematiche legate a tali strumenti attengono al loro grado di affidabilità poiché le risorse, quasi sempre, si concentrano sugli obiettivi che, in virtù delle indicazioni fornite dallo strumento, risultano essere più critici.

Il rischio, infatti, è di generare profilature severamente alterate degli individui. Così come ulteriori rischi di profilazione²³ si registrano con riguardo all’utilizzo dei mezzi predittivi in tema di recidiva.

¹⁹ Tra i più noti si ricorda *ChatGPT*.

²⁰ I. ANRÒ, *Intelligenza artificiale e giustizia*, intervento tenuto il 27.6.2025 a Bari in occasione della II *Summerschool AISDUE* dal titolo *La regolamentazione europea delle tecnologie digitali*.

²¹ Sul punto si richiamano le riflessioni di M. STRONATI, *The judicial decision between legal gaps and technological innovation: some suggestions from the 19th and 20th centuries*, in E. CALZOLAIO (a cura di), *La decisione nel prisma dell’intelligenza artificiale*, Padova, 2020, p. 37 ss.,

²² W.L. PERRY, B. MCINNS, C.C. PRICE, S.C. SMITH, J.S. HOLLYWOOD, *Predictive policing. The role of crime forecasting in law enforcement operations*, (2003) in www.rand.org.

²³ Tale preoccupazione è altresì espressa nel considerando n. 42 del Regolamento (UE) 1689/2024 secondo il quale «dovrebbero essere vietate le valutazioni del rischio effettuate in relazione a persone fisiche intese a determinare la probabilità che queste ultime commettano un reato o volte a prevedere il verificarsi di un reato

È noto, sul punto, quanto occorso nel caso statunitense *Loomis*²⁴, la cui fama è dovuta all'impiego dell'algoritmo denominato COMPAS il quale aveva condotto alla condanna, a sei anni di carcere, nei confronti di un cittadino americano laddove il giudice, all'atto di motivare la sua decisione, ha riportato, unitamente ad altre valutazioni, quanto prodotto dal *software* in questione circa il rischio di recidiva dell'imputato, identificato come soggetto altamente pericoloso²⁵.

E ciò sulla base dell'inserimento, da parte di chi aveva elaborato il programma, di dati che a priori penalizzavano l'elemento razziale, così da mettere in evidenza la possibilità di un maggiore rischio di commissione di delitti da parte di persone di carnagione scura.

Ne deriva che l'errore apparentemente commesso dalla macchina resta, per i motivi appena richiamati, comunque riconducibile all'uomo. Pertanto, sulla scorta di tale ultima riflessione, risulterebbe ancora piuttosto difficoltoso pensare di sostituire il giudice - essere umano con il giudice - robot, soprattutto se, attraverso quest'ultimo scenario, si immagina di soddisfare ampiamente l'esigenza di certezza²⁶.

Non risulta quindi difficile immaginare che le violazioni dei principi fondamentali, primo fra tutti il divieto di discriminazione, derivano anche dal fatto che i sistemi di IA non possono essere considerate entità neutre poiché producono risultati elaborati sulla base dei dati che vengono loro forniti e che rispecchiano, pertanto, le scelte dei programmatori²⁷.

Nei sistemi di intelligenza artificiale destinati ad essere applicati nel settore della giustizia, coloro i quali sono demandati all'addestramento della macchina assumono un ruolo centrale e sarebbe tuttavia auspicabile che gli stessi operassero con la costante supervisione e collaborazione di giuristi esperti della materia di riferimento; viceversa, si registrerebbe il rischio di rendere operativi degli strumenti dannosi, ingannevoli e comunque in severo contrasto con i valori e le tutele preposte al corretto svolgimento di un processo penale.

Il funzionamento di tali dispositivi si basa su un'indefinita quantità di dati che vengono messi in correlazione tra di loro. Da ciò deriva che tutti coloro che prendono parte alla vicenda processuale devono poter conoscere i dati che, sin dall'inizio, vengono inseriti all'interno dell'algoritmo²⁸.

Per ciò che attiene all'utilizzo di queste tecnologie ed il rapporto con il "giudicante umano" è opportuno muovere da alcune considerazioni di ordine generale. La correlazione di dati sopra accennata non trova spazio nella dimensione emulativa del pensiero e del ragionamento tipicamente umano. Invero, le innovazioni tecnologiche che ad oggi conosciamo risulterebbero incompatibili con le operazioni interpretative sul piano giuridico, laddove queste implicano una valutazione ed un costante bilanciamento i quali si fondano,

effettivo o potenziale unicamente sulla base della loro profilazione o della valutazione dei loro tratti della personalità e delle loro caratteristiche». Fra queste ultime il medesimo paragrafo richiama elementi «quali la cittadinanza, il luogo di nascita, il luogo di residenza, il numero di figli, il livello di indebitamento o il tipo di automobile».

²⁴ Corte Suprema del Wisconsin, *State v. Loomis* (881 N.W. 2d 749 (Wis 2016)).

²⁵ Sul fronte europeo, il caso spagnolo *VioGen* (2022) - quale algoritmo impiegato per prevedere il rischio di violenza di genere - che ha riguardato la giovane Lobna Hemid si colloca tra quelle operazioni predittive i cui esiti possono dar vita a forti criticità nella tutela dei diritti e delle libertà degli individui.

²⁶ M. LUCIANI, *La decisione giudiziaria robotica*, in *AIC*, n. 3/2018, p. 884 ss.

²⁷ R. E. KOSTORIS, *Intelligenza artificiale, strumenti predittivi e processo penale* in *disCrimen* dal 5.3.2024, <https://discrimen.it/>.

²⁸ L. PRESSACCO, *Intelligenza artificiale e ragionamento probatorio nel processo penale*, in G. DI PAOLO, L. PRESSACCO (a cura di), *Intelligenza artificiale e processo penale. Indagini, prove, giudizio*, in *Quaderni della Facoltà di Giurisprudenza dell'Università di Trento* vol. 63/2022, p. 91 ss.

nella fase specificamente decisionale del giudice, su aspetti di natura intuitiva e pacificamente discrezionale²⁹.

Attualmente, inoltre, gli strumenti predittivi operano secondo uno schema per casi che tende a ricercare, nelle risposte fornite in passato per determinati quesiti, le soluzioni a problemi ancora da risolvere.

In questa prospettiva, lo strumento predittivo nelle mani di un operatore che ben conosce gli orientamenti decisionali più datati potrebbe portare a ridurre eventuali errori. Invero, con riferimento alla fase decisoria, deve rammentarsi che l'assistenza eventualmente fornita è strutturata per orientarsi verso una continuità giurisprudenziale. Di tal guisa, un difensore a conoscenza del dato fornito all'algoritmo potrà scegliere quale strategia portare avanti poiché, confidando che nel sistema in cui opera vi sia una scarsa probabilità che il giudice si discosti dall'andamento fino a quel momento seguito, potrebbe essergli noto, già a monte, l'esito del giudizio instaurato³⁰.

Le violazioni, tanto sul piano discriminatorio quanto su quello della protezione dei dati³¹, la cui tutela è sancita dagli articoli 8 e 21 della Carta dei diritti fondamentali dell'Unione, rischiano di subire un incremento per via dell'impiego frequente di dati non specificamente determinati e vaghi.

Ci si chiede se tali meccanismi, basati su correlazioni tra enormi quantità di dati, possano farsi strada nella ricostruzione dei nessi di causalità tipici della fase decisoria, offrendo soluzioni, più o meno attendibili, sulle statuizioni future che il magistrato sarà chiamato a rendere in casi simili.

Su tale assunto sarebbe auspicabile orientarsi nella consapevolezza che tali strumenti, seppur in grado di elaborare soluzioni inedite per quesiti non previsti dalla macchina *ab origine*, non riescono a connettere "logicamente" i diversi assunti portati in evidenza dalle risultanze processuali. Si potrebbe, quindi, affermare che l'automazione, di per sé, osta allo sviluppo di un percorso causale così come non consente di operare un vero e proprio controllo sugli esiti prodotti tanto da non permettere di contestarli attraverso argomentazioni razionali.

Quanto sin qui esposto, con precipuo riferimento all'utilizzo degli algoritmi predittivi per finalità decisorie, renderebbe evidente il pregiudizio che si ripercuote, dal loro impiego, nel procedimento penale andando ad intaccare il principio del contraddittorio nella formazione della prova e, più ampiamente, i diritti di difesa così come sanciti dai testi sovranazionali posti a tutela dei diritti fondamentali nell'Unione europea³².

5. Il controllo umano dei processi operativi nel Regolamento (UE) 2024/1689

Per ciò che attiene ai settori definiti ad «alto rischio», il Regolamento in oggetto dispone di principi ed obblighi che sanciscono la progettazione antropocentrica più volte richiamata dal testo della normativa. L'articolo 14, infatti, delinea uno dei punti chiave della disciplina

²⁹ G. UBERTIS, *Intelligenza artificiale, giustizia penale, controllo umano significativo* in *Diritto penale contemporaneo* n. 4/2020, p. 82 ss.

³⁰ C. CASTELLI, D. PIANA, *Giustizia predittiva. La qualità della giustizia in due tempi*, in *Questione giustizia*, n. 2/2018.

³¹ Sul punto si richiama la disciplina dettata dal Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche in materia di trattamento dei dati personali e alla libera circolazione di tali dati, noto anche come *General Data Protection Regulation* (GDPR).

³² Il riferimento è al Titolo VI - inerente alla giustizia - della Carta dei diritti fondamentali dell'Unione europea (2016/C 202/02).

sull'intelligenza artificiale mediante la previsione di «strumenti di interfaccia uomo-macchina adeguati, in modo tale da poter essere efficacemente supervisionati da persone fisiche durante il periodo in cui sono in uso»³³.

La sorveglianza da parte dell'uomo consentirebbe, a colui il quale questa è affidata, di comprendere le potenzialità e le limitazioni, legate al funzionamento della macchina, e dedicarsi al superamento di eventuali distorsioni o malfunzionamenti. Questo permetterebbe altresì, in particolar modo all'interno dei sistemi impiegati nella produzione di informazioni finalizzate a decisioni che un soggetto dovrà prendere, di interpretare in maniera corretta l'*output* per poi eventualmente discostarsi dal prodotto del sistema e decidere di ignorarlo, *bypassarlo*, superarlo o, più semplicemente, non utilizzarlo³⁴.

Va inoltre evidenziato come lo stesso Regolamento preveda, all'art. 86, il diritto alla spiegazione dei singoli processi decisionali. La norma garantisce a chiunque sia destinatario di una pronuncia adottata dal *deployer*³⁵, sulla base di quanto elaborato da un sistema di IA ad alto rischio di cui all'allegato III del Regolamento - il cui contenuto incida in maniera significativa sull'individuo per avere un impatto negativo sulla sua salute, sicurezza o diritti fondamentali - il diritto di ottenere dei chiarimenti significativi sul ruolo che il sistema ha avuto nel corso della procedura decisionale posta in essere nonché sugli elementi principali della pronuncia resa³⁶.

Tuttavia è bene precisare, in via conclusiva, che non tutti i sistemi di IA impiegati nel campo della giustizia sono da considerare come ad «alto rischio»³⁷.

Lo stesso Regolamento, tenuto conto che l'apparato giudiziario consta anche di attività amministrative ed accessorie rispetto all'esercizio della funzione strettamente giurisdizionale, ha distinto delle modalità in cui l'impiego dell'intelligenza artificiale in tale settore può risultare utile per rendere celere e efficace la macchina giudiziaria. Tra queste attività, ad esempio, rientrerebbero l'organizzazione degli uffici, la raccolta dei documenti, la comunicazione e la gestione dell'attività dei cancellieri e così via.

L'impiego di strumenti tecnologici altamente evoluti per lo svolgimento di tali attività che, in quanto tali, non sono in grado di incidere sulle garanzie ed i valori fondamentali, risulterebbe comunque meritevole di un'adeguata considerazione dal momento che, il miglioramento dei sistemi giudiziari in termini di efficienza permette di perseguire, da questo punto di vista, gli obiettivi fissati dall'Unione europea³⁸.

In tale contesto, anche la Corte di giustizia dell'Unione europea ha sviluppato, nel tempo, una strategia in materia di intelligenza artificiale³⁹. Per il settore fin qui trattato, la Corte si è mostrata aperta all'utilizzo di nuove tecnologie nel campo della giustizia, soprattutto nei termini di un ampio miglioramento qualitativo, per ciò che attiene al ruolo

³³ Art. 14, rubricato *Sorveglianza umana*, del Regolamento (UE) 2024/1689.

³⁴ In ottica di controllo che il *deployer* è chiamato ad eseguire prima di utilizzare un sistema di IA, il Regolamento UE 2024/1689 pone all'articolo 27 una «Valutazione d'impatto sui diritti fondamentali per i sistemi di IA ad alto rischio».

³⁵ Dal considerando n. 13 dell'*AI Act* che definisce il *deployer* come «qualsiasi persona fisica o giuridica, compresi un'autorità pubblica, un'agenzia o altro organismo, che utilizza un sistema di IA sotto la sua autorità».

³⁶ G. CANZIO, *AI Act e processo penale: sfide e opportunità*, 14 ottobre 2024, <https://www.sistemapenale.it>.

³⁷ L'articolo 6 paragrafo 3 dell'*AI Act* prevede una deroga ai precedenti paragrafi che, viceversa, stabiliscono le condizioni per la qualificazione di un sistema di IA nella categoria «alto rischio».

³⁸ C. TERESI, *L'AI Act nell'ottica del processual-penalista: uno sguardo preliminare*, in *Rivista Penale Diritto e Procedura* n. 3/2024, p. 361 ss.

³⁹ Il documento è consultabile all'indirizzo web: curia.europa.eu/jcms/upload/docs/application/pdf/2023-11/cjeu_ai_strategy.pdf.

delle macchine nella funzione decisoria⁴⁰, e di una potenziale riduzione per un verso della durata dei giudizi, per l'altro del margine di errore nelle cause⁴¹.

Allo stesso tempo, vengono delineati i rischi e le preoccupazioni che deriverebbero dal coinvolgimento, in questa specifica fase, dell'IA. La Corte, infatti, nel ribadire che i nuovi strumenti digitali potrebbero effettivamente svolgere un ruolo di assistenza nei confronti dei magistrati, sottolinea l'impossibilità di giungere ad una completa automatizzazione delle decisioni. Di conseguenza afferma il pericolo concreto di decisioni troppo "meccaniche" comunque basate su modelli algoritmici che, per le loro caratteristiche, non sono in grado attualmente di tenere conto delle peculiarità, anche immateriali, e delle complessità che potrebbero registrarsi in un settore come quello giudiziario. Un eccessivo affidamento agli strumenti di intelligenza artificiale porterebbe, così, ad una riduzione della capacità discrezionale dei giudicanti trasformando il procedimento di un mero esercizio matematico probabilistico⁴².

Considerando, altresì, il divieto di adottare una decisione che si basi unicamente su trattamenti automatizzati⁴³, l'orizzonte tracciato anche dalla Corte di giustizia consentirebbe di propendere verso un approccio collaborativo tra l'uomo e la macchina nel rispetto dell'indipendenza degli organi della magistratura limitato ad una funzione meramente assistenziale nella fase decisionale che, in senso stretto, deve rimanere un'attività a guida umana.

Sempre in tema di indipendenza del potere giudiziario, non pare superfluo evidenziare che le decisioni prodotte mediante l'impiego di sistemi di IA potrebbero compromettere la fiducia che i cittadini ripongono nelle Autorità preposte alla tutela dei propri interessi, nonché i principi di *rule of law*, tra cui il diritto ad un giusto processo⁴⁴.

La disamina delle disposizioni previste dall'*AI Act*, per ciò che attiene al settore della giustizia, consente di rilevare il rapporto che intercorre tra una prospettiva strettamente giuridica e una prospettiva etica della disciplina sovranazionale.

Il modo in cui l'Unione ha scelto di disciplinare la materia, mediante la previsione di deroghe e formule ampie persino per le pratiche considerate vietate nonché per le disposizioni inerenti alle «attività di contrasto»⁴⁵, implica la necessità di un rinvio alla disciplina dettata dal diritto armonizzato dell'UE o, altresì, al diritto nazionale chiamato ad adeguarsi alla normativa sovranazionale.

Per ciò che attiene specificamente al procedimento penale, i principi che derivano tanto dalle tradizioni costituzionali, quanto dai testi sovranazionali in materia di giusto processo e tutela dei diritti fondamentali, esigono la previsione di uno schema di regole ben preciso che tenga conto di tutte le fasi procedurali, incluse quelle prodromiche alla sua instaurazione,

⁴⁰ Paragrafo n. 2, obiettivo n. 2, 2.1, 2.2, 2.3 della Strategia.

⁴¹ A. CORRERA, *Il ruolo dell'Intelligenza artificiale nel paradigma europeo dell'E-justice. Prime riflessioni alla luce dell'AI Act*, in *Quaderni AISDUE*, n. 2/2024.

⁴² S. A. BEVILACQUA, *La strategia dell'intelligenza artificiale della Corte di Giustizia dell'Unione europea in Q-DIGITO Quaderni di Organizzazione e Trasformazione Digitale della Giustizia*, n.1 2024, <https://www.ledijournals.com>, p. 123 ss.

⁴³ Art. 22 del Regolamento UE n. 2016/679.

⁴⁴ Comunicazione della Commissione, *Digitalizzazione della giustizia nell'Unione europea Un pacchetto di opportunità*, 2 dicembre 2020 COM(2020)710 final.

⁴⁵ Il riferimento è all'Allegato III, paragrafo 6, dell'*AI Act*.

affinché venga assicurato un equilibrio di poteri tra gli organi della magistratura e le parti private⁴⁶.

6. Conclusioni

L'influenza reciproca che il settore della giustizia ed i sistemi di IA riflettono l'uno sugli altri dà vita ad un modello ibrido che deve trovare spazio all'interno del quadro dei valori europei⁴⁷.

I principi contenuti nell'*AI Act*, da ricondursi in ottica generale a obiettivi di trasparenza, proporzionalità e sicurezza, necessitano di un'attuazione costante e concreta da parte degli Stati membri.

Per monitorare l'andamento della digitalizzazione nel settore della giustizia è stata pubblicata, ad inizio 2025, la Strategia europea in materia di giustizia elettronica per il periodo 2024-2028⁴⁸. Il piano d'azione delineato si concentra sugli interventi che gli Stati membri dovranno mettere in atto con riferimento a misure già adottate e da adottare a livello UE entro i cinque anni a seguire.

La Strategia, tra i cui obiettivi principali vi è il miglioramento dell'accesso alla giustizia e l'ottimizzazione dei sistemi giudiziari sotto il profilo dell'efficacia, pone al centro i diritti fondamentali e i valori dello Stato di diritto, principi cardine dell'ordinamento dell'Unione anche alla luce dell'articolo 2 del Trattato sull'Unione europea.

Dato l'enorme potenziale che gli strumenti di IA presentano nel panorama giudiziario europeo, l'Unione intende ribadire che l'obiettivo da perseguire nei prossimi anni attiene al perfezionamento di un approccio «digitale per default»⁴⁹ che punti, oltre ad uno snellimento delle procedure, anche un elevato livello di interoperabilità tra apparati diversi, al fine di superare i confini tra gli ordinamenti degli Stati membri.

Una novità di rilievo anticipata nella Strategia riguarda il passaggio da iniziative prettamente volontarie dei singoli Paesi a uno schema obbligatorio fondato su «atti legislativi che garantiscono una maggiore certezza del diritto»⁵⁰ poiché dotati di quella efficacia vincolante che punterebbe al raggiungimento di una disciplina omogenea e uniforme, sulla materia, in tutto il territorio dell'Unione.

Alla luce della normativa di riferimento e dei punti di forza delle nuove tecnologie che si intendono impiegare con maggiore efficacia nel settore della giustizia, ma soprattutto dei potenziali rischi più volte richiamati, è necessario che l'Unione attenzi l'evoluzione dei sistemi tecnologici e digitali posti a servizio dei cittadini e delle imprese presenti sul territorio e all'interno degli ordinamenti statali.

⁴⁶ Non meno rilevanti i momenti che attengono all'intervento dei sistemi di intelligenza artificiale per quel che riguarda la validità e la valutazione delle prove in giudizio, così come la necessità di garantire che tutte le parti possano partecipare al procedimento prima che il giudice fondi il suo convincimento su quanto emerso in dibattimento.

⁴⁷ N. LIPARI, *Il diritto del nuovo millennio tra giurisdizionalizzazione ed algoritmo*, in *Accademia - Rivista dell'Associazione Civilisti Italiani*, n. 4 del 30 Aprile 2024.

⁴⁸ Strategia europea in materia di giustizia elettronica 2024-2028, C/2025/437 in GUUE Serie C del 16 gennaio 2025.

⁴⁹ S. SIGNORIN, *Verso l'applicazione trasversale*, cit. p. 207

⁵⁰ Punto 15 par. A della strategia europea in materia di giustizia elettronica 2024-2028.

Sul punto, potrebbe risultare utile seguire l'andamento dell'intervento che la Commissione europea ha effettuato, con l'avvio di una consultazione pubblica sui sistemi di IA ad alto rischio, e che condurrà alla pubblicazione, entro il 2 febbraio 2026, delle linee guida sulle potenziali implementazioni dell'articolo 6 dell'*AI Act*⁵¹. La consultazione, conclusasi nel luglio 2025, consentirà alla Commissione di esaminare diversi esempi di utilizzo effettivo degli strumenti di IA, per i sistemi ad alto rischio, nonché di delineare le esigenze di carattere operativo dei diversi attori del settore.

Peraltro, è nel settore della giustizia penale in particolare che potrebbe risultare più sottile il confine tra un impiego proficuo dell'intelligenza artificiale e un utilizzo rischioso per le garanzie e le libertà che, in questo ambito, vengono costantemente in rilievo. L'approccio pratico non può discostarsi dal modello antropocentrico che continua a vedere l'uomo come il fulcro attorno al quale costruire una realtà digitale sempre più efficiente.

Le innovazioni in campo tecnologico non possono lasciare indietro le ragioni per cui le stesse vengono impiegate - vale a dire i bisogni della società odierna - e pregiudicare severamente i diritti fondamentali degli individui, specialmente quando questi incidono sulla materia penale.

La peculiarità dell'equilibrio auspicato richiede una valutazione costante degli effetti che il Regolamento (UE) 2024/1689 avrà in un futuro prossimo. A ciò si aggiunga che il buon esito dell'intervento legislativo potrà essere inoltre garantito dalle efficaci attività di controllo svolte dalle istituzioni di riferimento come la Commissione europea e la Corte di giustizia⁵² cui è demandato rispettivamente di vigilare sulla corretta applicazione e interpretazione della disciplina europea.

⁵¹ Il documento della consultazione è reperibile al sito web: <https://digital-strategy.ec.europa.eu/it/news/commission-launches-public-consultation-high-risk-ai-systems>.

⁵² A titolo meramente esemplificativo si fa riferimento ad una domanda di pronuncia pregiudiziale proposta dal Sąd Okręgowy w Warszawie (Polonia) il 26 febbraio 2025 con riferimento agli artt. 2 e 19 TUE e 20 e 47 CDFUE, Causa C-159/25, Rowicz, in <https://curia.europa.eu>.



LUIGI PIGNA*

IA GENERATIVA E TRATTAMENTO DEI DATI PERSONALI. LA TUTELA DEI DIRITTI FONDAMENTALI DELL'UTENTE ALLA LUCE DELLA RECENTE INIZIATIVA DI META PER L'ADDESTRAMENTO DELL'IA

SOMMARIO: 1. L'avvento dell'IA generativa. Cosa sono gli LLM e i connessi rischi per la privacy. Un'introduzione. – 2. I principi del RGPD come argine alle nuove tecnologie e il caso ChatGPT/GPDP. – 3. Meta alla ricerca del suo “posto al sole” nel settore dell'IA generativa. – 4. Segue. Il parere dell'EDPB nel segno della sentenza *Meta Platforms*. – 5. Riflessioni a margine dell'informativa di Meta – 6. Considerazioni conclusive

1. L'avvento dell'IA generativa. Cosa sono gli LLM e i connessi rischi per la privacy. Un'introduzione

L'attenzione pubblica attuale sull'intelligenza artificiale riguarda l'IA generativa, o *GAI*, che è specializzata nella creazione di nuovi contenuti, siano essi testi, voci, immagini o video¹. Un esempio di IA generativa è un *chatbot* basato su un modello linguistico di grandi dimensioni (*LLM*), come *ChatGPT*². Gli utenti interagiscono con questi sistemi inserendo un *prompt*, ovvero una domanda o richiesta, a cui l'IA risponde generando contenuti nuovi e pertinenti³. Questi modelli sono progettati per produrre *output* originali basati sui dati utilizzati durante la fase di addestramento. Una delle qualità più significative è la loro capacità di adattarsi a molteplici utilizzi, anche diversi da quelli previsti nella fase iniziale di

* Dottorando di ricerca in “Diritto pubblico, comparato e internazionale”, curriculum in “Ordine internazionale e diritti umani”, Università di Roma “La Sapienza”.

¹ Per un'analisi approfondita delle fasi in cui si articola la costruzione di un modello di intelligenza artificiale generativa si veda A. BANDI, P. ADAPA, Y. KUCHI, *The power of generative AI: A review of requirements, models, input-output formats, evaluation metrics, and challenges*, in *Future internet*, 2023.

² V. L. FLORIDI, *AI as agency without intelligence: On chatGPT, large language models, and other generative models*, in *Philosophy e technology*, 2023; AA.VV., *Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration*, in *Journal of information technology case and application research*, 2023.

³ Un *prompt* è un testo in linguaggio naturale che richiede all'IA generativa di eseguire una determinata attività. V. D. J. SOLOVE, *Artificial intelligence and privacy*, in *Florida Law Review*, 2025; C. GRIECO, *Il diritto d'autore nell'AI Act: Error 404 – page not found?*, in F. FERRI (a cura di), *L'Unione europea e la nuova disciplina sull'intelligenza artificiale: questioni e prospettive*, Rivista Quaderni AISDUE, Fascicolo speciale, 2/2024, Editoriale Scientifica, pp. 157-180.

addestramento e adattati a contesti specifici tramite tecniche di riaddestramento, integrazione o *fine-tuning*, basate su *set* di dati mirati e coerenti con le esigenze dell'utente o con uno specifico settore applicativo⁴. Tuttavia, sia nella fase di sviluppo sia nell'uso successivo, l'impiego dell'IA comporta il trattamento di dati personali, che nel tempo sono aumentati in quantità e varietà, con implicazioni rilevanti in termini di tutela della *privacy*.

Orbene, a partire dalla fine del 2022, con il successo di *ChatGPT* e di altri strumenti ad alte prestazioni, l'IA generativa ha conosciuto una crescita senza precedenti, affermandosi come protagonista non solo dell'innovazione tecnologica, ma anche del dibattito giuridico e regolatorio. Il termine “generativo” non è casuale, ma si riferisce ad un dominio dove le macchine diventano veri e propri creatori, dotati di un'abilità straordinaria, la capacità non solo di predire, ma di concepire e plasmare realtà del tutto nuove. Gli LLM, infatti, sono sistemi capaci di generare linguaggio naturale come *output*, a partire da un addestramento effettuato su enormi quantità di dati testuali⁵. Sebbene ogni modello di IA, in senso tecnico, possa essere considerato “generativo” in quanto produce risultati (come previsioni o regole), la nozione assume rilevanza specifica quando ci si riferisce a sistemi che generano informazioni nuove e non predefinite, intrinsecamente legate ai dati d'addestramento⁶. In questo contesto si inseriscono anche i modelli multimodali (*Large Multimodal Models, LMM*), come *GPT-4* o *LLaMA 4* di *Meta*, in grado di processare simultaneamente testo, immagini, suoni e video, ampliando in modo significativo le potenzialità applicative dell'IA⁷. Questi modelli si distinguono per la loro versatilità, configurandosi come strumenti di “utilizzo generale”, in grado di svolgere compiti eterogenei: dalla traduzione linguistica al riassunto testuale, dal supporto alla programmazione alla generazione creativa di contenuti, fino all'assistenza nella presa di decisioni complesse⁸.

La crescente autonomia e versatilità di tali modelli consente loro di cogliere strutture profonde nei dati e generare contenuti sempre più sofisticati, ma al contempo solleva interrogativi rilevanti sul piano della trasparenza, della responsabilità e, soprattutto, della tutela dei diritti fondamentali⁹. In effetti, gli LLM emergono come una risorsa preziosa, offrendo diversi vantaggi. È imperativo, tuttavia, riconoscere e comprenderne appieno le

⁴ DPC, *AI, Large Language Models and Data Protection*, <https://www.dataprotection.ie/en/dpc-guidance/blogs/AI-LLMs-and-Data-Protection> (6 luglio 2025).

⁵ Per un'analisi approfondita delle fasi in cui si articola la costruzione di un modello di intelligenza artificiale generativa si veda A. BANDI, P. ADAPA, Y. KUCHI, *The power of generative AI: A review of requirements, models, input-output formats, evaluation metrics, and challenges*, in *Future internet*, 2023, pp. 7-10.

⁶ V. D. FOSTER, *Generative deep learning*, O'Reilly Media, California, 2023, p. 5.

⁷ Per una disamina approfondita dei vari approcci definitivi si veda M. C. BUITEN, *Towards intelligent regulation of artificial intelligence*, in *European journal of risk regulation*, 2019, p. 41 ss.

⁸ In ambito sanitario, per esempio, l'impiego degli LLM nei sistemi di supporto alle decisioni cliniche ha aperto nuovi scenari: questi strumenti offrono raccomandazioni terapeutiche basate su evidenze e mostrano un potenziale significativo nel migliorare l'efficienza e la qualità dell'assistenza medica. È recente la notizia dello sviluppo da parte di Microsoft di un'IA in grado di identificare correttamente la diagnosi in fino all'85% dei casi clinici pubblicati sul *New England Journal of Medicine (NEJM)*, con un tasso di successo oltre quattro volte superiore rispetto a un gruppo di medici esperti. MAI-DxO raggiunge inoltre la diagnosi corretta in modo più efficiente dal punto di vista dei costi rispetto ai medici. V. Microsoft AI, *The Path to Medical Superintelligence*, <https://microsoft.ai/new/the-path-to-medical-superintelligence/> (16 luglio 2025).

⁹ V. M. L. MINSKY, *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind*, Simon e Schuster, New York, 2006, p. 95.

limitazioni. Sebbene possano costituire uno strumento efficace per la ricerca, le spiegazioni e l'ottimizzazione, esiste il rischio che le risposte siano fuorvianti o imprecise¹⁰.

È in questo contesto che l'Unione europea ha avviato una risposta normativa organica con il Regolamento sull'intelligenza artificiale (*Artificial Intelligence Act, AIA*), che si propone di disciplinare in modo trasversale l'intero ciclo di vita dei sistemi di IA, compresi quelli generativi¹¹. Tuttavia, per comprendere appieno la portata della disciplina, è necessario considerare anche altri strumenti normativi già esistenti nell'ordinamento dell'Ue, primo fra tutti il Regolamento generale sulla protezione dei dati (RGPD), il cui intreccio con l'IA generativa si sta rivelando sempre più complesso e strategico¹².

Una delle questioni più controverse riguarda, infatti, il trattamento dei dati personali nei processi di addestramento e utilizzo dei modelli di intelligenza artificiale. Problemi quali la memorizzazione involontaria di dati sensibili, le tecniche di *model inversion* (che permettono la ricostruzione di dati individuali a partire dai parametri del modello)¹³, e la gestione dei dati inseriti nei *prompt* dagli utenti, pongono sfide significative rispetto ai principi fondamentali del RGPD: liceità del trattamento, trasparenza, limitazione della finalità, minimizzazione e tutela dei soggetti vulnerabili, come i minori.

In particolare, si discute della correttezza della base giuridica utilizzata per il trattamento dei dati a fini di addestramento. Emblematico, in tal senso, è il recente caso *Meta*, che ha annunciato l'intenzione di utilizzare, a partire da maggio 2025, i contenuti pubblici condivisi dagli utenti dell'Ue per addestrare i propri modelli di IA generativa, facendo leva sulla base giuridica del legittimo interesse (art. 6, par. 1, lett. f, RGPD). Tale iniziativa ha fatto seguito a un precedente stop dovuto alle obiezioni sollevate da autorità di regolazione e attivisti per la *privacy*¹⁴. In risposta, la *Data Protection Commission (DPC)* irlandese ha richiesto un parere formale al Comitato europeo per la protezione dei dati (*EDPB*) ai sensi dell'art. 64 del RGPD. Il parere, pubblicato nel dicembre 2024, ha indicato una serie di criteri che le autorità nazionali devono tenere in considerazione nella valutazione della legittimità di questi trattamenti¹⁵.

¹⁰ D. GANGULI ET AL., *Predictability and surprise in large generative models*, in *ACM Conference on Fairness, Accountability, and Transparency*, 2022.

¹¹ Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale), in GU L, 2024/1689, 12.7.2024.

¹² 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), in GU L 119, 4.5.2016, p. 1–88.

¹³ Le tecniche di *model inversion* (letteralmente, inversione del modello) sono metodi attraverso i quali, analizzando i parametri interni di un modello di intelligenza artificiale già addestrato, è possibile ricostruire informazioni sensibili o personali presenti nei dati su cui è stato addestrato. Più semplicemente, se ChatGPT non memorizza direttamente tutti i dati con cui è stato addestrato, potrebbe comunque contenere nei suoi parametri delle tracce di quei dati. Le tecniche di *model inversion* cercano di estrarre o dedurre queste informazioni. Relativamente alle tecniche di *model inversion*, si veda AA.VV., *Deep Learning Model Inversion Attacks and Defenses: A Comprehensive Survey*, consultabile al sito <https://arxiv.org/pdf/2501.18934>.

¹⁴ Uno su tutti, l'associazione noyb di Maximilian Schrems. V. *nobyb urges 11 DPAs to immediately stop Meta's abuse of personal data for AI*, consultabile al sito <https://nobyb.eu/en/nobyb-urges-11-dpas-immediately-stop-metas-abuse-personal-data-ai>.

¹⁵ V. EDPB, Parere 28/2024 su taluni aspetti relativi alla protezione dei dati ai fini del trattamento dei dati personali nel contesto dei modelli di IA, adottato il 17 dicembre 2024.

Questo contributo si propone di analizzare il rapporto tra intelligenza artificiale generativa e protezione dei dati personali, con un focus specifico sulle implicazioni normative del reg. 2016/679 e sulle recenti evoluzioni interpretative fornite dall'EDPB. Particolare attenzione sarà dedicata alla valutazione della base giuridica del legittimo interesse, ai margini di applicazione del consenso, e alla necessità di una *governance* trasparente e bilanciata del ciclo di vita dei modelli generativi. Il tutto alla luce della coesistenza, e talvolta tensione, tra il regolamento sull'IA e il quadro giuridico sulla protezione dei dati.

2. I principi del RGPD come argine alle nuove tecnologie e il caso ChatGPT/GPDP.

La protezione dei dati personali rappresenta una delle sfide più rilevanti nello sviluppo e nell'impiego dell'intelligenza artificiale generativa. Il corretto funzionamento dei sistemi di intelligenza artificiale richiede la disponibilità di ingenti quantità di dati¹⁶. Un passaggio cruciale in questo senso è proprio quello della raccolta dei dati che derivano dall'impiego di dati personali nel *training*, dalla possibile memorizzazione involontaria delle informazioni, dall'inserimento di dati personali nei prompt da parte degli utenti e, infine, dai contenuti generati dagli stessi modelli. Al riguardo, il nuovo regolamento sull'intelligenza artificiale non si occupa della raccolta e del trattamento dei dati, i quali rimangono pertanto soggetti agli strumenti vigenti e, in particolare, al regolamento 2016/679, sulla protezione dei dati personali¹⁷.

È possibile dunque individuare diverse aree problematiche che vanno dalla definizione di una base giuridica appropriata per l'addestramento; al diritto alla cancellazione (*right to erasure*), dalla tutela dei minori alla più generale gestione dei dati sensibili, fino al principio di limitazione delle finalità e di minimizzazione dei dati¹⁸.

Tra i principi fondamentali previsti dal RGPD, particolare rilievo assumono quelli della minimizzazione¹⁹ e del consenso dell'interessato²⁰. Il principio di minimizzazione prevede la raccolta unicamente dei dati necessari al raggiungimento della finalità del trattamento. Esso si fonda sull'idea che trattare solo i dati strettamente necessari riduca il rischio di violazioni della *privacy* o fughe di informazioni²¹. Il requisito del consenso prevede invece che l'interessato debba manifestare espressamente la sua volontà che i dati vengano trattati per una o più specifiche finalità. Il consenso deve essere, oltre che inequivocabile, anche libero, specifico e informato²².

¹⁶ European Parliament Research Service, *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*, 2020, consultabile online al sito [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU\(2020\)641530_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf).

¹⁷ V. G. CONTALDI, *Intelligenza artificiale e dati personali*, in *Ordine Internazionale e Diritti Umani*, 2021.

¹⁸ V. C. NOVELLI, F. CASOLARI, P. HACKER, G. SPEDICATO, L. FLORIDI, *Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity*, in *Computer Law & Security Review*, 2024.

¹⁹ Art. 5 reg. 2016/679.

²⁰ Art. 6, par. 1, lett. a) e art. 7 reg. 2016/679.

²¹ V. AA.VV., *The Data Minimization Principle in Machine Learning*, in *FACt '25: Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 2025; H. VAN KOLFSCHOOTEN, *EU regulation of artificial intelligence: Challenges for patients' rights*, in *Common Market Law Review*, 2022, pp. 100 ss.

²² Sull'argomento cfr. A. VIVARELLI, *Il consenso al trattamento dei dati personali nell'era digitale. Sfide tecnologiche e soluzioni giuridiche*, Napoli, 2019; C. KOOLEN, *Transparency and Consent in Data-Driven Smart Environment*, in *European Data Protection Law Review*, 2021; Y. POULLET, *Consentement et Rgpd: des zones d'ombre!*, in *Droit de la consommation*, 2019.

Entrambi questi principi appaiono, in verità, poco rispondenti al funzionamento concreto dei sistemi di intelligenza artificiale. Tali sistemi, infatti, utilizzano una vasta quantità di dati. Del resto, maggiore è la disponibilità e l'accuratezza delle informazioni utilizzate, più performanti risultano le previsioni indirizzate agli utenti²³. Allo stesso tempo, i sistemi di IA generativa utilizzano dati collazionati con finalità diverse per elaborare i propri modelli predittivi. Una delle manifestazioni più rilevanti dei *LLM* consiste proprio nella loro versatilità come strumento di utilizzo generale per una vasta gamma di compiti, compresi quelli non specificamente contemplati durante l'addestramento²⁴.

Parallelamente, per i sistemi di IA generativa, l'aggregazione automatica dei dati su larga scala può complicare il rispetto dei principi del consenso²⁵. La base giuridica più rilevante nel RGPD è il consenso. Tuttavia, nel caso di *set* di dati di grandi dimensioni, che includono informazioni personali provenienti da un numero elevato di persone sconosciute agli sviluppatori, ottenere un consenso valido da ciascun individuo non è realisticamente possibile a causa degli elevati costi di transazione²⁶. Inoltre, l'utilizzo di *LLM* basati su dati raccolti dal web e impiegati per scopi non sempre prevedibili è difficilmente compatibile con un consenso informato e specifico²⁷. Inoltre, è sostanzialmente accertato che la maggior parte degli utenti tende a prestare poca attenzione nel momento in cui presta il proprio assenso. Nell'attuale ecosistema digitale, il consenso ha progressivamente perso efficacia. Di fronte al potere delle grandi piattaforme, risulta evidente come gli utenti difficilmente prestino un consenso davvero libero e informato²⁸.

L'aspetto della base giuridica per le attività di trattamento dei dati resta tuttavia una questione fondamentale. Qualsiasi trattamento di dati personali, comprese operazioni come la raccolta, la memorizzazione o la diffusione, deve fondarsi su una base giuridica ai sensi dell'art. 6 del RGPD. I modelli generativi vengono addestrati su grandi moli di dati, spesso tratti da fonti online accessibili pubblicamente, che possono contenere informazioni personali. Senza un fondamento giuridico specifico, tale trattamento risulterebbe illecito. Seguendo l'approccio adottato dal legislatore nel regolamento generale sulla tutela dei dati personali, il consenso costituisce solo uno dei fondamenti giuridici del trattamento, assumendo così la stessa valenza delle altre condizioni possibili di liceità²⁹. Tale flessibilità consente di prendere in considerazione i presupposti di legittimità più appropriati sia per il titolare che per l'interessato. Se però è vero che tutte queste condizioni presentano analoga valenza (eccetto, stando al considerando 46, la residualità del trattamento dei dati necessario a proteggere un interesse vitale dell'interessato o di terzi), va tenuto presente che il ventaglio dei diritti dell'interessato si diversifica rispetto ad esse. In questo modo, il diritto alla

²³ V. T.Z. ZARSKY, *Incompatible: The GDPR in the Age of Big Data*, in *Seton Hall Law Review*, 2017.

²⁴ Cfr. AA.VV., *Is chatGPT a general-purpose natural language processing task solver?*, in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023.

²⁵ V. EDPB, Linee guida 5/2020 sul consenso ai sensi del regolamento (UE) 2016/679, adottate il 4 maggio 2020.

²⁶ M. MOURBY, K.O. CATHAOIR, C.B. COLLIN, *Transparency of machine-learning in healthcare: the GDPR & European health law*, in *Computer Law Security Review*, 2021.

²⁷ Al riguardo, vedi il parere della *Autoriteit Persoonsgegevens*, *scraping bijna altijd illegal*, 1 May 2024, consultabile al sito <https://www.autoriteitpersoonsgegevens.nl/actueel/ap-scraping-bijna-altijd-illegaal>.

²⁸ Sul consenso nel mercato digitale, v. F. BATTAGLIA, *Il consumatore digitale nel diritto dell'Unione europea*, Napoli, 2023. Si veda anche D.J. SOLOVE, *Murky consent: An approach to the fictions of consent in privacy law*, in *Boston university law review*, 2024.

²⁹ In particolare: art. 6, lett b) esecuzione di un contratto; c) adempimento di un obbligo legale; d) salvaguardia di un interesse vitale dell'interessato; e) tutela di un interesse pubblico; f) legittimo interesse del titolare o di terzi, del reg. 2016/679.

portabilità³⁰ opera solo quando il trattamento si basi sul consenso o sia necessario per l'esecuzione di un contratto, mentre il diritto di opposizione³¹, salvo il trattamento diretto alle finalità di *marketing*, riguarda solo il trattamento compiuto ai sensi dell'art. 6, par. 1, lett. e) o f)³².

Per ragioni sia giuridiche che economiche (ossia i costi legati alla raccolta del consenso), la formazione degli *LLM* può quindi basarsi solo sul cosiddetto *test* di bilanciamento previsto dall'art. 6, par. 1, lett. f) RGPD³³. Secondo tale disposizione, è lecito trattare i dati personali se ciò è “necessario per il perseguimento del legittimo interesse del titolare del trattamento” (cioè l'ente che sviluppa il modello), purché tali interessi non prevalgano sui diritti e le libertà fondamentali degli interessati (ossia le persone i cui dati vengono utilizzati).

Tuttavia, l'ipotesi del legittimo interesse del titolare o di terzi merita particolare attenzione, in quanto rappresenta una delle condizioni di liceità più complesse da applicare³⁴. Essa si basa, innanzitutto, su un concetto indeterminato (non è prevista una definizione all'interno del regolamento) e richiede un bilanciamento tra l'interesse del titolare del trattamento e i diritti e le libertà fondamentali dell'interessato, in particolare se si tratta di un minore. Sia il RGPD che la Corte di giustizia hanno riconosciuto espressamente che diversi interessi possono costituire un interesse legittimo ai fini del trattamento dei dati personali³⁵. È il titolare a dover valutare autonomamente la prevalenza del proprio interesse, assumendosi la responsabilità di tale scelta, in linea con il principio di *accountability*³⁶. Allo stesso tempo, questa base giuridica non può diventare né una scorciatoia per eludere la richiesta di consenso, né una clausola inapplicabile per eccesso di rigidità. Il RGPD, unitamente al ruolo del Comitato e del Garante europeo per la protezione dei dati, ha fornito criteri per guidare i titolari in questa valutazione, che deve essere concreta, documentata e rispettosa dei diritti degli interessati³⁷.

Orbene, stabilire se il *test* di bilanciamento costituisca una base giuridica valida richiede una valutazione caso per caso³⁸. In generale, quando l'intelligenza artificiale viene utilizzata per scopi socialmente utili o positivi³⁹, la posizione del titolare del trattamento dei dati è vista in modo più favorevole sotto il profilo giuridico; inoltre, se l'interessato può ragionevolmente aspettarsi che i propri dati vengano utilizzati per l'addestramento di un'IA, questo rafforza la

³⁰ Art. 20 reg. 2016/679.

³¹ Art. 21 reg. 2016/679.

³² V. D. POLETTI, *Le condizioni di liceità del trattamento dei dati personali*, in *Giurisprudenza italiana*, 2019.

³³ V. F. ZUIDERVEEN BORGESIU, S. KRUIKEMEIER, S. BOERMAN, N. HELBERGER, *Tracking walls, take-it-or-leave-it choices, the GDPR, and the ePrivacy regulation*, in *European Data Protection Law Review*, 2018, p. 353-368.

³⁴ V. C.A. TSCHIDER, *AI's Legitimate Interest: Towards a public benefit privacy model*, in *Houston journal of health law & policy*, 2021, p. 178 ss.

³⁵ V. EDPB, *Guidelines 1/2024 on processing of personal data based on Article 6(1)(f) GDPR*, adopted on 8 October 2024, *cit.*, p. 7.

³⁶ Art. 5, par. 2 reg. 2016/679: «2. Il titolare del trattamento è competente per il rispetto del paragrafo 1 e in grado di provarlo («responsabilizzazione»).

³⁷ Considerando 47, 48 e 49 reg. 2016/679; cfr. EDPB, *Guidelines 1/2024*.

³⁸ V. E. GIL GONZÁLEZ, P. DE HERT, *Understanding the legal provisions that allow processing and profiling of personal data—an analysis of GDPR provisions and principles*, in *ERA Forum* 19, 2019, pp. 597-621; M. DONNELLY, M. McDONAGH, *Health Research, Consent and the GDPR Exemption*, in *European Journal of Health Law*, 2019, pp. 97-119.

³⁹ Ad esempio, per migliorare l'assistenza sanitaria, ridurre le disuguaglianze, ottimizzare l'accesso ai servizi pubblici o promuovere la sostenibilità ambientale.

posizione del titolare⁴⁰. Tuttavia, questa condizione è difficilmente soddisfatta nella pratica. Non a caso, l'adozione di strategie mirate a rafforzare la tutela della *privacy*, come la pseudonimizzazione, la trasparenza nei confronti degli interessati o l'utilizzo di tecniche di crittografia⁴¹, può contribuire in modo significativo a rendere legittimo l'addestramento dei modelli di intelligenza artificiale quando questo si fonda sulla base giuridica del legittimo interesse del titolare del trattamento. Allo stesso tempo, la legittimità del trattamento può essere esclusa qualora emergano fattori critici, come la natura e l'estensione del trattamento, la tipologia dei dati coinvolti, soprattutto se si tratta di dati sensibili, un insufficiente livello di trasparenza o un grado limitato di controllo esercitabile dalle persone interessate. In tali casi, il bilanciamento potrebbe dare esito negativo, impedendo al titolare di invocare il legittimo interesse come base giuridica per l'addestramento del modello.

Ulteriore questione, legata all'uso di ampi dataset per l'addestramento dei modelli di IA, riguarda il coinvolgimento di categorie particolari di dati personali, il cui trattamento è generalmente vietato ai sensi dell'art. 9 del RGPD. Tuttavia, alla luce di una recente sentenza della Corte di giustizia, tali informazioni non devono necessariamente riferirsi direttamente a categorie protette, come origine etnica o razziale, religione, età o salute, per rientrare nell'ambito dell'art. 9. È sufficiente che il trattamento dei dati consenta di rivelare informazioni rientranti in una di queste categorie⁴². Ciò che conta è la capacità del titolare del

⁴⁰ Così il considerando 47 del reg. 2016/679: «I legittimi interessi di un titolare del trattamento, compresi quelli di un titolare del trattamento a cui i dati personali possono essere comunicati, o di terzi possono costituire una base giuridica del trattamento, a condizione che non prevalgano gli interessi o i diritti e le libertà fondamentali dell'interessato, tenuto conto delle ragionevoli aspettative nutrite dall'interessato in base alla sua relazione con il titolare del trattamento. Ad esempio, potrebbero sussistere tali legittimi interessi quando esista una relazione pertinente e appropriata tra l'interessato e il titolare del trattamento, ad esempio quando l'interessato è un cliente o è alle dipendenze del titolare del trattamento. In ogni caso, l'esistenza di legittimi interessi richiede un'attenta valutazione anche in merito all'eventualità che l'interessato, al momento e nell'ambito della raccolta dei dati personali, possa ragionevolmente attendersi che abbia luogo un trattamento a tal fine. Gli interessi e i diritti fondamentali dell'interessato potrebbero in particolare prevalere sugli interessi del titolare del trattamento qualora i dati personali siano trattati in circostanze in cui gli interessati non possano ragionevolmente attendersi un ulteriore trattamento dei dati personali. Posto che spetta al legislatore prevedere per legge la base giuridica che autorizza le autorità pubbliche a trattare i dati personali, la base giuridica per un legittimo interesse del titolare del trattamento non dovrebbe valere per il trattamento effettuato dalle autorità pubbliche nell'esecuzione dei loro compiti. Costituisce parimenti legittimo interesse del titolare del trattamento interessato trattare dati personali strettamente necessari a fini di prevenzione delle frodi. Può essere considerato legittimo interesse trattare dati personali per finalità di marketing diretto».

⁴¹ Cfr. art. 25 reg. 2016/679: «1. Tenendo conto dello stato dell'arte e dei costi di attuazione, nonché della natura, dell'ambito di applicazione, del contesto e delle finalità del trattamento, come anche dei rischi aventi probabilità e gravità diverse per i diritti e le libertà delle persone fisiche costituiti dal trattamento, sia al momento di determinare i mezzi del trattamento sia all'atto del trattamento stesso il titolare del trattamento mette in atto misure tecniche e organizzative adeguate, quali la pseudonimizzazione, volte ad attuare in modo efficace i principi di protezione dei dati, quali la minimizzazione, e a integrare nel trattamento le necessarie garanzie al fine di soddisfare i requisiti del presente regolamento e tutelare i diritti degli interessati».

⁴² V. la sentenza della Corte del 4 luglio 2023, causa C-252/21, *Meta Platforms Inc. e a. c. Bundeskartellamt* (Condizioni generali di utilizzo di un social network) ECLI:EU:C:2023:537, punto 73: «Alla luce di quanto precede, occorre rispondere alla seconda questione, lettera a), dichiarando che l'articolo 9, paragrafo 1, del RGPD deve essere interpretato nel senso che, nel caso in cui un utente di un social network online consulti siti Internet oppure applicazioni correlati a una o più delle categorie menzionate da tale disposizione e, se del caso, inserisca in essi dati, iscrivendosi oppure effettuando ordini online, il trattamento di dati personali da parte dell'operatore di tale social network online – consistente nel raccogliere, tramite interfacce integrate, cookie o simili tecnologie di registrazione, dati risultanti dalla consultazione di tali siti e di tali applicazioni nonché i dati inseriti dall'utente, nel mettere in relazione l'insieme di tali dati con l'account del social network di quest'ultimo e nell'utilizzare detti dati – deve essere considerato un «trattamento di categorie particolari di dati personali» ai

trattamento di dedurre tratti sensibili dai dati disponibili, indipendentemente dall'intento dell'operatore.

Pertanto, gli sviluppatori di modelli di IA, dinanzi alla possibilità di dedurre dati sensibili in fase di addestramento o deployment, e rientrare con il trattamento, nell'ambito dell'art. 9, dovranno ricorrere a una delle deroghe previste al par. 2 della medesima disposizione. Tuttavia, al di fuori del consenso esplicito, tali eccezioni sono raramente disponibili⁴³.

Sul diritto all'oblio, previsto dall'art. 17 del RGPD, i modelli linguistici di grandi dimensioni possono memorizzare dati specifici del *set* di addestramento, migliorando le prestazioni su distribuzioni sbilanciate, ma esponendo gravi rischi per la *privacy*. Attacchi come il *membership inference* e il *training data extraction* dimostrano che è possibile risalire a informazioni sensibili, anche solo interrogando il modello⁴⁴. Inoltre, anche se i produttori di

sensi di detta disposizione, il quale è in linea di principio vietato, fatte salve le deroghe previste da detto articolo 9, paragrafo 2, qualora tale trattamento di dati sia tale da rivelare informazioni rientranti in una di dette categorie, a prescindere dal fatto che tali informazioni riguardino un utente di tale social network o qualsiasi altra persona fisica».

⁴³ Art. 9, par. 2 reg. 2016/679: «Il paragrafo 1 non si applica se si verifica uno dei seguenti casi: a) l'interessato ha prestato il proprio consenso esplicito al trattamento di tali dati personali per una o più finalità specifiche, salvo nei casi in cui il diritto dell'Unione o degli Stati membri dispone che l'interessato non possa revocare il divieto di cui al paragrafo 1; b) il trattamento è necessario per assolvere gli obblighi ed esercitare i diritti specifici del titolare del trattamento o dell'interessato in materia di diritto del lavoro e della sicurezza sociale e protezione sociale, nella misura in cui sia autorizzato dal diritto dell'Unione o degli Stati membri o da un contratto collettivo ai sensi del diritto degli Stati membri, in presenza di garanzie appropriate per i diritti fondamentali e gli interessi dell'interessato; c) il trattamento è necessario per tutelare un interesse vitale dell'interessato o di un'altra persona fisica qualora l'interessato si trovi nell'incapacità fisica o giuridica di prestare il proprio consenso; d) il trattamento è effettuato, nell'ambito delle sue legittime attività e con adeguate garanzie, da una fondazione, associazione o altro organismo senza scopo di lucro che persegue finalità politiche, filosofiche, religiose o sindacali, a condizione che il trattamento riguardi unicamente i membri, gli ex membri o le persone che hanno regolari contatti con la fondazione, l'associazione o l'organismo a motivo delle sue finalità e che i dati personali non siano comunicati all'esterno senza il consenso dell'interessato; e) il trattamento riguarda dati personali resi manifestamente pubblici dall'interessato; f) il trattamento è necessario per accertare, esercitare o difendere un diritto in sede giudiziaria o ogniqualvolta le autorità giurisdizionali esercitano le loro funzioni giurisdizionali; g) il trattamento è necessario per motivi di interesse pubblico rilevante sulla base del diritto dell'Unione o degli Stati membri, che deve essere proporzionato alla finalità perseguita, rispettare l'essenza del diritto alla protezione dei dati e prevedere misure appropriate e specifiche per tutelare i diritti fondamentali e gli interessi dell'interessato; h) il trattamento è necessario per finalità di medicina preventiva o di medicina del lavoro, valutazione della capacità lavorativa del dipendente, diagnosi, assistenza o terapia sanitaria o sociale ovvero gestione dei sistemi e servizi sanitari o sociali sulla base del diritto dell'Unione o degli Stati membri o conformemente al contratto con un professionista della sanità, fatte salve le condizioni e le garanzie di cui al paragrafo 3; i) il trattamento è necessario per motivi di interesse pubblico nel settore della sanità pubblica, quali la protezione da gravi minacce per la salute a carattere transfrontaliero o la garanzia di parametri elevati di qualità e sicurezza dell'assistenza sanitaria e dei medicinali e dei dispositivi medici, sulla base del diritto dell'Unione o degli Stati membri che prevede misure appropriate e specifiche per tutelare i diritti e le libertà dell'interessato, in particolare il segreto professionale; j) il trattamento è necessario a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici in conformità dell'articolo 89, paragrafo 1, sulla base del diritto dell'Unione o nazionale, che è proporzionato alla finalità perseguita, rispetta l'essenza del diritto alla protezione dei dati e prevede misure appropriate e specifiche per tutelare i diritti fondamentali e gli interessi dell'interessato».

⁴⁴ Ritornando all'esempio dell'ambito medico, gli *LLM* raccolgono una vasta quantità di dati, come possono essere quelli relativi ad una determinata malattia. Se in un dataset ci sono pochi dati riguardanti questa specifica malattia, il modello di IA risponderebbe ugualmente ai prompt/domande inseriti dagli utenti (quale ad esempio i sintomi legati alla stessa), ma questo non attraverso una generalizzazione, ma perché ricorda direttamente il contenuto originale di quei pochi dati a disposizione sempre relativi a quella malattia. Al riguardo v. F.

LLM, come OpenAI, dichiarano di rispettare il diritto alla cancellazione, non è chiaro come ciò sia possibile, poiché le informazioni personali possono essere contenute in più forme all'interno di un LLM, rendendo molto complesso identificare ed eliminare punti specifici, soprattutto quando i dati non sono strutturati, come possono essere i numeri di telefono. Inoltre, le richieste di rimozione avviate da un singolo interessato possono risultare insufficienti, in particolare nei casi in cui informazioni identiche siano state diffuse da più utenti durante l'interazione con il modello⁴⁵. Detto in altri termini, il diritto alla cancellazione applicato ai modelli di GAI non pare rappresentare una garanzia adeguata a tutela della *privacy*, avendo gli stessi la capacità di inferire ugualmente dettagli di informazioni dai dati originali anche una volta eliminati⁴⁶.

I modelli di *Generative AI* hanno sollevato preoccupazioni anche riguardo al consenso dei minori riguardo ai contenuti adatti all'età, vista la possibilità di generare output non adatti ai minori⁴⁷.

Emblematico è stato il caso del Garante per la protezione dei dati personali italiano, che nel 2023 ha disposto un blocco temporaneo di ChatGPT⁴⁸. Le principali criticità individuate riguardavano l'assenza di una base giuridica valida per la raccolta dei dati personali, una scarsa trasparenza nel trattamento dei dati utilizzati per l'addestramento del modello e la mancanza di sistemi efficaci per impedire l'accesso ai minori, con il rischio che questi fossero esposti a contenuti non adeguati alla loro età e maturità⁴⁹.

In risposta, OpenAI ha sospeso temporaneamente l'accesso al servizio in Italia, introducendo misure correttive tra cui un sistema per consentire agli utenti di richiedere la cancellazione dei propri dati personali ai sensi del RGPD e un modulo per la verifica dell'età durante la registrazione. Dopo l'adozione di queste modifiche e l'adeguamento delle *policy*, il GPDP ha revocato il blocco ad aprile 2023⁵⁰.

Nel frattempo, l'EDPB, il Comitato europeo per la protezione dei dati, ha avviato una *task force* per monitorare a livello europeo la conformità dei modelli generativi come ChatGPT al quadro normativo. Tra i nodi più delicati, vi è la legittimità del trattamento dei dati personali, che deve sempre basarsi su una delle condizioni previste dall'art. 6 RGPD e, in caso di dati sensibili, anche su una delle eccezioni indicate all'art. 9.

Una delle principali aree critiche ha riguardato la raccolta dei dati tramite *web scraping*, ovvero l'estrazione automatizzata di informazioni da fonti pubblicamente accessibili online. Questi dati possono includere anche elementi sensibili, per cui l'adozione di misure di salvaguardia risulta essenziale. OpenAI ha invocato come base giuridica il legittimo interesse, ma ciò, come visto, richiede un attento bilanciamento tra gli interessi dell'azienda e i diritti e le libertà

DERNONCOURT, J. Y. LEE, O. UZUNER, P. SZOLOVITS, *De-identification of patient notes with recurrent neural networks*, in *Journal of the American Medical Informatics Association*, 2017, pp. 596–606.

⁴⁵ Al riguardo, v. AA.VV., *What Does it Mean for a Language Model to Preserve Privacy?*, in *FACIT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022, pp. 2282 ss.

⁴⁶ E. DE CRISTOFARO, *An overview of privacy in machine learning*, 2020, consultabile al sito <https://arxiv.org/pdf/2005.08679v1>.

⁴⁷ Ai sensi dell'art. 8, par. 2 del reg. 2016/679, il titolare del trattamento deve adottare «sforzi ragionevoli per verificare [...] che il consenso dei minori sia prestato o autorizzato da chi esercita la responsabilità genitoriale, tenendo conto della tecnologia disponibile».

⁴⁸ Su tutte le questioni relative all'utilizzo di ChatGPT, si veda K. NAGHIYEV, *ChatGPT From a Data Protection Perspective*, in *Baku State University Law Review*, 2024.

⁴⁹ V. F. GUALDI, A. CORDELLA, *Theorizing the regulation of generative AI: lessons learned from Italy's ban on ChatGPT*, in *Proceedings of the 57th Hawaii International Conference on System Sciences*, 2024.

⁵⁰ Si rimanda a Garante per la protezione dei dati personali, *ChatGPT: OpenAI riapre la piattaforma in Italia garantendo più trasparenza e più diritti a utenti e non utenti europei*, consultabile al sito <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9881490>.

degli interessati. È stato inoltre chiarito che il semplice fatto che un'informazione sia pubblicamente accessibile non equivale a un consenso esplicito dell'interessato alla diffusione dei suoi dati sensibili, come richiesto dall'art. 9, par. 2, lett. e) RGPD.

Per mitigare i rischi, sono state suggerite misure tecniche e organizzative come l'anonimizzazione preventiva dei dati, l'esclusione di alcune fonti (ad esempio i profili *social* pubblici), e l'impiego di filtri per evitare la raccolta di dati appartenenti a categorie particolarmente protette. Anche i contenuti generati dagli utenti, come i *prompt* inseriti e gli *output* del sistema, sono oggetto di trattamento da parte di OpenAI, che ha dichiarato di utilizzarli per migliorare le proprie tecnologie. Anche in questo caso è stato invocato il legittimo interesse, ma resta fondamentale che gli utenti siano pienamente informati di tale utilizzo e messi nella condizione di esercitare un'opposizione effettiva⁵¹.

Nel gennaio 2024, il GPDP ha notificato un nuovo provvedimento correttivo e sanzionatorio nei confronti di OpenAI, concludendo l'istruttoria avviata l'anno precedente⁵². Secondo l'Autorità, l'azienda ha trattato i dati personali degli utenti senza aver stabilito una base giuridica adeguata, ha violato i principi di trasparenza previsti dal RGPD e non ha segnalato una violazione dei dati avvenuta nel marzo 2023. Inoltre, ha omesso di adottare misure per la verifica dell'età, esponendo potenzialmente i minori a contenuti inadatti.

Per rimediare, e utilizzando per la prima volta i nuovi poteri previsti dall'articolo 166 del Codice *Privacy*⁵³, il Garante ha ordinato a OpenAI di realizzare una campagna di comunicazione istituzionale della durata di sei mesi su media tradizionali e digitali. Lo scopo è quello di informare il pubblico sul funzionamento di *ChatGPT*, sul trattamento dei dati, anche dei non utenti, e sui diritti esercitabili, come opposizione, rettifica e cancellazione.

Come ulteriore conseguenza, il Garante ha imposto a *OpenAI* una sanzione amministrativa di 15 milioni di euro, valutata anche alla luce della collaborazione dimostrata dalla società durante il procedimento⁵⁴. Infine, poiché OpenAI ha stabilito la propria sede europea in Irlanda, il GPDP ha trasmesso il fascicolo all'autorità irlandese, che diventa ora l'autorità capofila competente per le eventuali violazioni ancora in corso, in base al meccanismo dello sportello unico previsto dal RGPD⁵⁵. Tale vicenda diventa rilevante alla luce delle implicazioni che può avere sul caso oggetto della presente trattazione.

3. Meta alla ricerca del suo "posto al sole" nel settore dell'IA generativa.

Passando alla recente iniziativa di Meta, l'obiettivo dichiarato è quello di potenziare le capacità di Meta *AI*, l'assistente basato su intelligenza artificiale integrato nei servizi di

⁵¹ EDPB, *Report of the work undertaken by the ChatGPT Taskforce*, 23 May 2024, consultabile al sito https://www.edpb.europa.eu/system/files/2024-05/edpb_20240523_report_chatgpt_taskforce_en.pdf

⁵² Anche in seguito al parere dell'EDPB del 18 dicembre 2024 richiesto dalla DPC irlandese sul caso Meta e che sarà oggetto di esame nei capitoli a seguire.

⁵³ L'articolo 166 del Codice della Privacy (Decreto Legislativo n. 196/2003), come modificato dal D.Lgs. 101/2018, disciplina i criteri per l'applicazione delle sanzioni amministrative pecuniarie e il procedimento per l'adozione dei provvedimenti correttivi e sanzionatori, raccordandosi con le disposizioni del reg. 2016/679, in particolare l'art. 83.

⁵⁴ Si rimanda al GPDP, *Comunicato stampa - ChatGPT, il Garante privacy chiude l'istruttoria. OpenAI dovrà realizzare una campagna informativa di sei mesi e pagare una sanzione di 15 milioni di euro*, consultabile al sito <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/10085432>.

⁵⁵ Considerando 127 reg. 2016/679.

Facebook, Instagram e WhatsApp. A tal fine, la società prevede di utilizzare contenuti pubblici condivisi da utenti adulti su Facebook e Instagram, come *post*, didascalie, commenti e *reel*. Con oltre 500 milioni di utenti mensili nell'Unione europea, Meta dispone di un patrimonio informativo vasto e variegato sia dal punto di vista linguistico che demografico, che intende valorizzare per rafforzare la propria posizione in un settore sempre più competitivo⁵⁶. Meta potrà inoltre raccogliere dati da fonti pubblicamente accessibili *online*, anche attraverso tecniche di *web scraping*, o ottenere informazioni fornite da soggetti terzi. Restano esclusi da tale attività i messaggi privati (protetti da crittografia end-to-end e soggetti a misure di trattamento riservato)⁵⁷, così come i contenuti pubblici provenienti da *account* appartenenti a utenti di età inferiore ai 18 anni.

Su quest'ultimo punto, l'informativa sulla *privacy* di Meta chiarisce che, nel caso in cui un utente pubblichi foto, *tag* o informazioni riguardanti un minore (ad esempio, un genitore che condivide immagini del proprio figlio) tramite un *account* pubblico, tali dati potrebbero essere utilizzati da Meta per addestrare i suoi sistemi di intelligenza artificiale⁵⁸. In tal caso, il minore, non essendo titolare di un *account* né avendo accettato alcuna informativa, non ha alcun controllo né possibilità di opporsi al trattamento. Lo stesso vale per gli adulti che non dispongono di un *account* o non utilizzano i servizi di Meta: anche loro possono essere coinvolti inconsapevolmente, qualora vengano menzionati o ritratti in contenuti pubblici pubblicati da altri utenti.

Ai sensi dell'art. 6 del RGPD, ogni trattamento di dati personali deve basarsi su una valida base giuridica. Meta fa riferimento all'interesse legittimo come fondamento per l'uso dei dati personali degli utenti, sostenendo che lo sviluppo e il miglioramento dell'intelligenza artificiale siano utili sia per sé stessa, sia per gli utenti, per soggetti terzi e per il pubblico in generale. L'obiettivo è fornire strumenti efficaci, sicuri e affidabili, favorendo l'accesso all'informazione, l'innovazione e la condivisione, anche attraverso la collaborazione con sviluppatori e ricercatori⁵⁹.

Considerando tale fondamento, gli utenti hanno il diritto di opporsi a questo tipo di trattamento ai sensi dell'art. 21 del RGPD. Per questo Meta ha previsto un meccanismo di *opt-out* per chi non vuole che i propri dati siano utilizzati in questo modo; tale diritto si estende anche ai non utenti, i cui dati personali pubblici o forniti da terzi potrebbero essere inclusi nel *dataset* per l'addestramento dell'IA⁶⁰.

Orbene, questa *policy*, come già accennato, è stata oggetto di accertamenti da parte della *Data Protection Commission* irlandese e ha incontrato opposizioni in Germania, dove la *Verbraucherzentrale NRW* (organizzazione tedesca per i diritti dei consumatori) ha richiesto un blocco urgente del trattamento. Tuttavia, il Tribunale Superiore Regionale di Colonia ha

⁵⁶ European Commission, *Supervision of the designated very large online platforms and search engines under DSA*, consultabile al sito <https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses#ecl-inpage-metaplatforms>.

⁵⁷ Meta, *About Private Processing*, consultabile al sito https://faq.whatsapp.com/2089630958184255?fbclid=IwY2xjawLQ1I5leHRuA2FlbQlXMAbicmlkETfVeGtwQ0Z5bjFyQXFzb3hFAR7uabUsXBkKrwYNrtaagv2tHFqovmV9PwH1tEXWbF8Zh-SYdHr23wT4V8GiBQ_aem_t4Bd-9D05f0PX_e_1pOP0Mg.

⁵⁸ Meta, Centro sulla privacy, *In che modo Meta utilizza le informazioni per i modelli e le funzioni dell'IA generativa*, consultabile al sito <https://www.facebook.com/privacy/genai>.

⁵⁹ Meta, Centro sulla privacy, *Normativa sulla privacy, Che cos'è l'Informativa sulla privacy e cosa contempla?*, consultabile al sito <https://www.facebook.com/privacy/policy>.

⁶⁰ Facebook, Centro Assistenza, *Opponiti all'uso delle tue informazioni per l'IA in Meta*, consultabile al sito https://www.facebook.com/help/contact/6359191084165019?entrypoint=privacy_center.

respinto tale richiesta, riconoscendo preliminarmente la legittimità del trattamento basato sull'interesse legittimo di Meta, ai sensi dell'art. 6, par. 1, lett. f) del RGPD, ritenendo che gli obiettivi perseguiti non possano essere raggiunti con misure meno invasive. Inoltre, le autorità di controllo, tra cui il Garante di Amburgo, hanno confermato che il trattamento di dati pubblicamente accessibili, anche tramite motori di ricerca, è conforme alla normativa, anche quando si tratta di dati sensibili o riferiti a terzi, inclusi i minori, grazie alle garanzie rafforzate adottate da Meta, come le tecniche di de-identificazione, il filtraggio dei dati e i sistemi di controllo sui risultati prodotti⁶¹. Il trattamento rimarrà comunque sotto stretta supervisione della DPC, che ha richiesto a Meta un report sull'efficacia delle misure implementate, previsto per ottobre 2025⁶².

4. *Segue. Il parere dell'EDPB nel segno della sentenza Meta Platforms*

In questo contesto si inserisce il parere dell'EDPB, richiesto dall'autorità irlandese per la protezione dei dati nel caso Meta, che ha offerto chiarimenti interpretativi sul trattamento dei dati personali ai fini dell'addestramento dei modelli di intelligenza artificiale. Il Comitato si è pronunciato su temi centrali come l'anonimizzazione, l'interesse legittimo e i confini del trattamento lecito, facendo riferimento a un consolidato orientamento giurisprudenziale della Corte di giustizia. Con specifico riguardo all'interesse legittimo, che rappresenta la base giuridica invocata da Meta, il titolare è tenuto a effettuare una valutazione articolata in tre fasi per determinare la liceità del trattamento dei dati personali. L'EDPB cita, in particolare, la decisione della Corte nella causa *Meta Platforms*. In quella pronuncia, la Corte aveva chiarito quanto già ribadito in una precedente sentenza, per cui il ricorso alla base giuridica del legittimo interesse, prevista dall'art. 6, par. 1, lett. f) del RGPD, richiede la presenza congiunta di tre condizioni: innanzitutto, deve esistere un interesse legittimo effettivamente perseguito dal titolare o da un terzo; in secondo luogo, il trattamento deve essere strettamente necessario al raggiungimento di tale finalità; infine, è necessario che gli interessi, i diritti o le libertà fondamentali della persona interessata non prevalgano su quelli del titolare⁶³.

Con riferimento al *test* di necessità, finalizzato a indagare se il trattamento sia effettivamente e realmente necessario per il perseguimento dell'interesse del titolare del trattamento, sempre la Corte, nel già citato caso *Meta Platforms*, ha stabilito che «spetta al giudice del rinvio verificare che il legittimo interesse al trattamento dei dati perseguito non possa ragionevolmente essere raggiunto in modo altrettanto efficace mediante altri mezzi meno pregiudizievoli per i diritti fondamentali degli interessati»⁶⁴. Condizione di necessità che andrebbe esaminata unitamente al principio della minimizzazione dei dati⁶⁵. Inoltre, tra le considerazioni necessarie da tener conto in sede di *test* di necessità, vi sono la possibilità di informare previamente gli utenti e chiedere a questi ultimi se desiderano che i loro dati siano

⁶¹ Oberlandesgericht Köln, *Meta darf Daten aus öffentlich gestellten Nutzerprofilen für KI-Training verwenden*, https://www.olg-koeln.nrw.de/beh_toppmeldungen_archiv/Default-Toppmeldung-2/index.php (13 luglio 2025).

⁶² DPC, *DPC statement on Meta AI*, 21st May 2025, consultabile al sito <https://www.dataprotection.ie/en/news-media/latest-news/dpc-statement-meta-ai>.

⁶³ V. sentenza della Corte (Seconda Sezione) del 4 maggio 2017, causa C-13/16, *Valsts policijas Rīgas reģiona pārvaldes Kārtības policijas pārvalde c. Rīgas pašvaldības SLA "Rīgas satiksme"*, ECLI:EU:C:2017:336, punto 28.

⁶⁴ *Meta Platforms*, punto 108.

⁶⁵ *Ibidem*, punto 109.

trattati per determinate finalità, il che consente loro di «conservare il controllo sulla divulgazione dei loro dati personali e di limitare così la divulgazione di tali dati a quanto è effettivamente necessario e pertinente rispetto alle finalità per le quali detti dati sono trasmessi e trattati»⁶⁶ e l'attuazione di garanzie tecniche, come l'anonimizzazione, che limitano comunque la facilità con cui gli interessati possono essere identificati⁶⁷.

Relativamente all'esercizio del bilanciamento, che consiste «nell'identificare e descrivere i diversi diritti e interessi contrapposti in gioco, vale a dire da un lato gli interessi, i diritti fondamentali e le libertà degli interessati, e dall'altro gli interessi del titolare del trattamento o di terzi»⁶⁸, la valutazione deve tenere in considerazione le circostanze specifiche del singolo caso⁶⁹. È infatti fondamentale valutare ogni volta quali siano le legittime aspettative della persona interessata, così come l'entità e le modalità del trattamento dei dati in questione. In particolare, la Corte, nel caso *Meta Platforms*, ha rilevato che, sebbene il «miglioramento del prodotto» non possa essere escluso, in linea di principio, come legittimo interesse, ha altresì osservato che «appare dubbio che [...] l'obiettivo diretto al miglioramento del prodotto possa – tenuto conto della portata di tale trattamento e del suo notevole impatto sull'utente, nonché della circostanza che quest'ultimo non possa ragionevolmente attendersi che tali dati siano trattati [...] – prevalere sui diritti fondamentali e sugli interessi di detto utente, tanto più nel caso in cui quest'ultimo sia minorenne»⁷⁰. È inoltre necessario considerare la gravità dei rischi potenziali per gli interessati. A tal fine, possono essere valutati diversi aspetti, come le modalità del trattamento (ad esempio, se i dati personali vengono incrociati con altri *set* di dati), l'estensione e la quantità del trattamento (ossia il volume totale dei dati, la quantità di dati per ciascun individuo, il numero complessivo degli interessati), la condizione dell'interessato (ad esempio, se si tratta di minori o soggetti vulnerabili), nonché la natura del rapporto tra l'interessato e il titolare del trattamento (ad esempio, se si tratta di un cliente). Un caso esemplificativo è rappresentato dall'uso del *web scraping* durante le fasi di sviluppo: se mancano adeguate garanzie, questa pratica può comportare effetti rilevanti sulle persone, a causa dell'elevata quantità di dati raccolti, del grande numero di soggetti coinvolti e della raccolta indiscriminata di informazioni personali⁷¹.

L'EDPB rileva che alcune questioni rilevanti in materia di protezione dei dati e intelligenza artificiale non sono state approfondite nel Parere, poiché escluse dall'ambito della richiesta formulata dall'Autorità irlandese. Tra queste rientra il trattamento dei dati particolari ai sensi dell'art. 9 RGPD, per il quale si dispone il divieto generale, salvo che i dati siano stati resi pubblici dall'interessato in modo esplicito e inequivocabile, attraverso un atto che manifesti in modo chiaro e volontario la volontà dell'interessato stesso⁷². Sempre l'EDPB ricorda il disposto di cui alla sentenza della Corte di giustizia nel caso *Meta Platforms*, secondo la quale, qualora un *dataset* contenga dati particolari non suscettibili di separazione da altre categorie di dati, il relativo trattamento deve essere qualificato come trattamento di dati particolari⁷³.

⁶⁶ V. sentenza della Corte del 4 ottobre 2024, causa C-621/22, *Koninklijke Nederlandse Lawn Tennisbond*, ECLI:EU:C:2024:857, punti 51-53.

⁶⁷ EDPB, Parere 28/2024, punto 75.

⁶⁸ EDPB, Parere 28/2024, punto 76.

⁶⁹ G. PROIETTI, *Algoritmi e interesse del titolare del trattamento nella circolazione dei dati*, in *Contratto e impresa*, 2022, p. 891.

⁷⁰ *Meta Platforms*, (punto 123).

⁷¹ V. EDPB, Parere 28/2024, punto 86.

⁷² Art. 9, par. 2 re. 2016/679.

⁷³ *Meta Platforms*, punto 73.

In conclusione, anche se l'esistenza di tali vantaggi può portare un'autorità di controllo, come la Commissione irlandese, a concludere che sugli interessi del titolare del trattamento o di un terzo non prevalgono gli interessi, i diritti e le libertà fondamentali degli interessati, tale conclusione può essere tratta solo a seguito di un'analisi effettuata caso per caso che tenga conto di tutti questi fattori.

5. *Riflessioni a margine dell'informativa di Meta*

Il quadro giuridico dell'Unione europea in materia di addestramento dell'intelligenza artificiale, in particolare dei modelli linguistici di grandi dimensioni (*LLM*), a partire da dati personali, si è delineato attraverso pareri delle autorità garanti e pronunce della Corte di giustizia. Tuttavia, casi recenti come quello di Meta, evidenziano come persistano significative incertezze sulla legittimità di tali pratiche. Si ritiene dunque necessaria una disamina specifica dell'informativa privacy di Meta, anche alla luce delle considerazioni fatte sul caso *ChatGPT*.

È proprio il ricorso all'interesse legittimo, su cui si basa l'informativa di Meta, a sollevare le maggiori criticità. La questione, tuttavia, si presenta particolarmente complessa. Si consideri che se nella sentenza *Meta Platforms*, più volte citata, la Corte di giustizia ha stabilito che Meta non può fare affidamento sull'art. 6, par. 1, lett. f) del RGPD per la pubblicità personalizzata all'interno di una piattaforma, anche perché gli interessati non si aspetterebbero ragionevolmente che i loro dati vengano usati a tali fini. Eppure, essendo le piattaforme Meta finanziate prevalentemente tramite pubblicità, tale trattamento risulterebbe sicuramente più prevedibile rispetto all'addestramento dell'IA, una tecnologia del tutto nuova e sconosciuta per gli utenti iscritti.

Si può senz'altro discutere come l'addestramento dell'intelligenza artificiale possa rispettare il principio di limitazione della finalità o di minimizzazione dei dati. Questi principi, sanciti rispettivamente dall'art. 5, par. 1, lett. b) e c) del RGPD, trovano un'esplicita conferma anche nell'*AI Act*, in particolare nei considerando 69 e 71, che delineano un approccio *human-centric* al trattamento dei dati. In particolare, il considerando 69 ribadisce la necessità di garantire la protezione dei dati lungo l'intero ciclo di vita dei sistemi di intelligenza artificiale, attraverso l'applicazione sistematica dei principi di minimizzazione e della protezione dei dati fin dalla progettazione e per impostazione predefinita⁷⁴. A sua volta, il considerando 71 sottolinea il ruolo essenziale della documentazione tecnica come strumento di trasparenza, volto a rendere verificabili le scelte progettuali e a tutelare i diritti fondamentali degli interessati. In tale contesto normativo, la conformità dell'addestramento degli algoritmi a questi principi non può essere presunta, ma richiede un'attenta valutazione *ex ante* delle finalità perseguite e della proporzionalità dei dati utilizzati⁷⁵.

Relativamente alla distinzione tra categorie di dati, l'informativa di Meta ammette che i contenuti utilizzati per l'addestramento dell'IA possono riguardare anche informazioni sensibili, come contatti personali o opinioni su temi delicati. Questo solleverebbe rilevanti questioni giuridiche, in particolare rispetto alle categorie particolari di dati personali⁷⁶, tra cui

⁷⁴ Art. 25 reg. 2016/679.

⁷⁵ V. B. O. OZPARLAK, E. METIN, *Designing Trust and Privacy: The Role of Legal Design in Ensuring Compliance with the EU AI Act*, in A. MOALLEM (eds), *HCI for Cybersecurity, Privacy and Trust*, 2025, p. 116 ss.

⁷⁶ Art. 9 reg. 2016/679.

orientamento sessuale, convinzioni religiose o politiche, frequentemente condivise su piattaforme come Facebook e Instagram. Il trattamento di tali dati è in linea di principio vietato, salvo specifiche eccezioni, tra cui quella prevista dall'art. 9, par. 2, lett. e), quando è l'interessato stesso a renderli manifestamente pubblici. Tuttavia, in assenza di impostazioni consapevoli o di un'adeguata informazione, non si può presumere che questi dati siano liberamente riutilizzabili, se non previo consenso esplicito. Sul punto, la Corte di giustizia si è espressa in modo chiaro nella sentenza *Schrems c. Meta Platforms Ireland Ltd*⁷⁷, chiarendo che un contenuto può essere considerato “manifestamente pubblico” solo se l'utente ha consapevolmente configurato i parametri di visibilità affinché sia accessibile a un numero illimitato di persone. In altre parole, è necessario che vi sia stato un consenso esplicito alla pubblicazione in senso ampio⁷⁸.

L'informativa solleva inoltre dubbi sul rispetto del principio di trasparenza e correttezza, soprattutto quando nei contenuti trattati compaiono soggetti terzi o minori, che non sono stati informati e non hanno strumenti per opporsi. L'RGPD richiede infatti una tutela rafforzata per i dati dei minori nel contesto digitale⁷⁹, e il semplice carattere pubblico di un contenuto non basta a legittimarne l'uso: occorre sempre una base giuridica adeguata. In questo contesto, il ricorso all'interesse legittimo da parte di Meta appare particolarmente controverso, soprattutto quando coinvolge soggetti vulnerabili⁸⁰.

In merito al sistema di *opt-out*, ossia al diritto di opporsi al trattamento, sussistono diverse criticità. Molti utenti, infatti, potrebbero non comprendere pienamente cosa implichi la qualificazione dei propri contenuti come “pubblici” nell'ambito dell'addestramento algoritmico, oppure potrebbero non accorgersi delle notifiche che li informano in merito. Inoltre, una volta che i dati sono stati acquisiti nei sistemi di addestramento, non sembra esistere una reale possibilità di rimuoverli. A ciò si aggiunge il fatto che, in alcuni casi, le informazioni personali potrebbero comunque essere trattate nonostante l'opposizione espressa, come ad esempio quando l'utente compare in un'immagine condivisa pubblicamente da altri, o viene menzionato in un *post* o in una didascalia pubblicata da terzi (oltre al dettaglio, non così irrilevante, degli effetti non retroattivi del diritto di opposizione)⁸¹. Si tratta di situazioni che toccano direttamente la tematica del diritto all'oblio, come delineato nella nota sentenza *Google Spain*⁸².

Oltre a queste considerazioni legate al RGPD, è necessario il riferimento all'art. 5, par. 1, lett. b) e c) del *Digital Markets Act* (DMA), per cui il legislatore dell'Unione stabilisce limiti molto precisi all'uso dei dati personali da parte dei *gatekeeper*⁸³. In particolare, Meta non

⁷⁷ V. sentenza della Corte del 4 ottobre 2024, causa C-446/21, *Maximilian Schrems c. Meta Platforms Ireland Limited*, ECLI:EU:C:2024:834, punto 83.

⁷⁸ Si veda anche L. PIGNA, *Schrems c. Meta. La Corte di giustizia si pronuncia sui limiti al trattamento dei dati personali per fini di pubblicità mirata*, in *Ordine Internazionale e Diritti Umani*, 2024.

⁷⁹ Considerando 38 e art. 8 reg. 2016/679.

⁸⁰ Sul trattamento di categorie particolari di dati, v. A. BUSACCA, *Le “categorie particolari di dati” ex art. 9 GDPR. Divieti, eccezioni e limiti alle attività di trattamento*, in *Ordine Internazionale e Diritti Umani*, 2018.

⁸¹ Facebook, Centro assistenza, *Opponiti all'uso delle tue informazioni per l'IA in Meta*, consultabile al sito https://www.facebook.com/help/contact/6359191084165019?entrypoint=privacy_center.

⁸² V. sentenza della Corte del 13 maggio 2014, causa C-131/12, *Google Spain*, ECLI:EU:C:2014:317. In tema di diritto all'oblio, cfr. S. MARTINELLI, *Diritto all'oblio e motori di ricerca: il bilanciamento tra memoria e oblio in internet e le problematiche poste dalla de-indicizzazione*, in *Il diritto dell'informazione e dell'informatica*, 2017; M. TAMPIERI, *Il diritto all'oblio e la tutela dei dati personali*, in *Responsabilità civile e previdenza*, 2017.

⁸³ Regolamento (Ue) 2022/1925 del parlamento europeo e del consiglio del 14 settembre 2022 relativo a mercati equi e contendibili nel settore digitale e che modifica le direttive (UE) 2019/1937 e (UE) 2020/1828 (regolamento sui mercati digitali), in GU L 265, 12.10.2022, pagg. 1–66.

potrebbe combinare dati da servizi diversi, come dati presi da Facebook con quelli raccolti da Instagram, WhatsApp, o da servizi esterni (es. un sito che usa il *login* di Facebook), a meno che non abbia il consenso esplicito dell'utente, così come non potrebbe trasferire e usare in modo incrociato dati personali da un servizio all'altro del medesimo *gatekeeper*⁸⁴.

6. Considerazioni conclusive

Lo sviluppo e l'addestramento di sistemi di intelligenza artificiale comportano necessariamente il trattamento di grandi quantità di dati (anche personali). Il paradosso è che la progettazione, la sperimentazione e la validazione di sistemi di IA affidabili e rispettosi dei diritti fondamentali richiedono l'accesso a dati personali rappresentativi e attendibili, sia sotto il profilo quantitativo che qualitativo. Ne consegue una tensione tra le esigenze tecniche dello sviluppo dell'IA e la normativa sulla protezione dei dati personali, specialmente per quanto riguarda la liceità della raccolta e del trattamento dei dati destinati alla costruzione dei *dataset* per l'addestramento di modelli di IA e di apprendimento automatico⁸⁵.

Il trattamento di qualsiasi informazione relativa a una persona fisica identificata o identificabile è soggetto al rispetto delle norme in materia di protezione dei dati. Pertanto, al momento della raccolta e dell'utilizzo di dati personali, il titolare del trattamento deve individuare una valida base giuridica, che può consistere nel consenso dell'interessato oppure in un'altra delle condizioni previste dal RGPD.

Considerate le difficoltà pratiche legate all'ottenimento di un consenso realmente informato e significativo nel contesto dello sviluppo dell'IA, e alla sua idoneità come base giuridica, l'interesse legittimo perseguito dal titolare o da un terzo emerge come una possibile alternativa. Molti titolari del trattamento, come le piattaforme digitali, tendono a fondare il trattamento dei dati personali sull'interesse legittimo. Tale interesse viene spesso formulato in termini di sviluppo tecnologico e innovazione, con riferimento alla creazione, fornitura e miglioramento di tecnologie basate sull'IA. Si richiama anche un presunto interesse di terzi, come sviluppatori o partner commerciali, e, più in generale, dell'interesse del pubblico nel beneficiare di tali tecnologie. A ciò si aggiunge un ulteriore livello giustificativo, legato alla sicurezza e all'affidabilità dei sistemi, nel tentativo di dimostrare che il trattamento è volto anche a tutelare gli stessi utenti o l'ecosistema digitale⁸⁶.

Tuttavia, il recente parere 28/2024 del Comitato Europeo per la Protezione dei Dati, che, come visto, ha ripreso una consolidata giurisprudenza della Corte in materia e i risultati della *task-force* per *ChatGPT*, ha chiarito ancora una volta alcuni punti fondamentali che mettono in discussione la piena sostenibilità di queste formulazioni. Il ricorso all'interesse legittimo come base giuridica per l'addestramento dei modelli di IA non può essere invocato in modo generico o strumentale. Esso deve fondarsi su un interesse concreto e specifico del titolare, e che deve essere effettivamente bilanciato con i diritti e le libertà fondamentali degli interessati. Non sono sufficienti, a tal fine, motivazioni di ordine economico o tecnologico, né l'invocazione di interessi pubblici generici. L'uso di tale base giuridica comporta un onere

⁸⁴ Cfr. R. PODSZUN, *Should Gatekeepers Be Allowed to Combine Data? Ideas for Art. 5(a) of the Draft Digital Markets Act*, in *GRUR International*, 2022.

⁸⁵ V. P. T. KRAMCSÁK, *Can legitimate interest be an appropriate lawful basis for processing Artificial Intelligence training datasets?*, in *Computer Law and Security Review*, 2023, *cit.*

⁸⁶ Meta, *Che cos'è l'Informativa sulla privacy e cosa contempla?*.

significativo per il titolare, che deve dimostrare la necessità del trattamento, adottare misure tecniche e organizzative adeguate, come la minimizzazione dei dati e la trasparenza, e garantire un livello di tutela coerente con le aspettative legittime degli interessati. Lo stesso regolamento sull'IA, richiamando esplicitamente il principio della *data protection by design* e *by default*⁸⁷, nonché l'obbligo di documentazione tecnica trasparente per i sistemi ad alto rischio, conferma che la protezione dei dati personali deve essere garantita lungo l'intero ciclo di vita dei sistemi di IA⁸⁸. Inoltre, come previsto dal considerando 67 del nuovo testo legislativo, il principio di trasparenza gioca anch'esso un ruolo chiave, che deve essere garantito con riguardo alla finalità originaria della raccolta dei dati utilizzati per l'addestramento. Tale protezione implica, tra l'altro, la minimizzazione dei dati trattati, in linea con quanto previsto dall'art. 5, par. 1, lett. c) e dall'art. 25 RGPD. Proprio questo principio di "progettazione preventiva" esclude approcci meramente reattivi o *ex post*, come tentativi di mitigare i rischi attraverso il filtraggio degli *output* o la limitazione delle risposte del sistema.

L'invocazione dell'interesse legittimo come base giuridica per il trattamento dei dati personali nel contesto dell'intelligenza artificiale non può dunque essere interpretata come un espediente per semplificare gli obblighi del titolare: al contrario, essa comporta un rafforzamento delle sue responsabilità. E l'interpretazione data dall'EDPB ha escluso, in definitiva, ogni approccio formalistico, ribadendo come la protezione dei dati personali non può essere subordinata agli obiettivi commerciali o di innovazione tecnologica.

Proprio perché caratterizzato da una certa elasticità applicativa, il ricorso a tale fondamento richiede un'analisi rigorosa e caso-specifica delle finalità perseguite, della proporzionalità del trattamento e dell'effettiva necessità dei dati trattati rispetto allo scopo dichiarato. Anche se, come visto nel caso dell'iniziativa di Meta, i dubbi vanno oltre la scelta della base giuridica del trattamento. A sollevare preoccupazione è soprattutto l'impatto dell'addestramento sui dati cosiddetti "particolari", specie quelli riguardanti minori. Su questo punto, l'*AI Act* dispone una tutela rafforzata di tali categorie di informazioni personali, prevista dall'art. 10, par. 5, che oltre ad essere soggette a limitazioni tecniche, nonché a misure più avanzate di sicurezza, le stesse non possono essere trasmesse, trasferite o consultate da terzi. I titolari del trattamento devono inoltre raccogliere solo i dati personali necessari e pertinenti a una finalità specifica⁸⁹. Anche l'*AI Act* riflette questo principio, richiedendo una valutazione della quantità e adeguatezza dei dati⁹⁰.

Inoltre, con l'aumento della dimensione e complessità dei *dataset* utilizzati dai fornitori di *LLM*, diventa sempre più importante potenziare le misure per la protezione dei dati personali contro possibili violazioni⁹¹. I metodi tradizionali, come la pseudonimizzazione e la cifratura, restano validi, ma possono non essere sufficienti da soli. Una nuova frontiera è

⁸⁷ Considerando 69 reg. 2024/1689.

⁸⁸ Considerando 71 reg. 2024/1689.

⁸⁹ Rispettivamente, art. 5, lett. b) e c) del reg. 2016/679: «I dati personali sono: b) raccolti per finalità determinate, esplicite e legittime, e successivamente trattati in modo che non sia incompatibile con tali finalità; un ulteriore trattamento dei dati personali a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici non è, conformemente all'articolo 89, paragrafo 1, considerato incompatibile con le finalità iniziali («limitazione della finalità»); c) adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità per le quali sono trattati («minimizzazione dei dati»).

⁹⁰ Art. 10, par. 2, lett. e) del reg. 2024/1689: «I set di dati di addestramento, convalida e prova sono soggetti a pratiche di governance e gestione dei dati adeguate alla finalità prevista del sistema di IA ad alto rischio. Tali pratiche riguardano in particolare: e) una valutazione della disponibilità, della quantità e dell'adeguatezza dei set di dati necessari».

⁹¹ Il riferimento dell'*AI Act* a questi metodi (articolo 10(5)(b)) ne sottolinea l'importanza,

quella del *machine unlearning*, una tecnica che consente di eliminare ogni traccia dell'informazione indesiderata all'interno del modello stesso⁹². Anche l'addestramento su banche dati più ridotte potrebbe ridurre eventuali rischi per la privacy degli utenti. Tuttavia, tali limitazioni avrebbero come effetto collaterale quello di ridurre la flessibilità e il potenziale innovativo dei modelli. In alternativa, può essere più efficace rafforzare le misure di protezione della *privacy* proporzionalmente alla grandezza del *dataset*⁹³.

In conclusione, gli strumenti per uno sviluppo delle nuove tecnologie, come la *GAI*, orientato al rispetto dei diritti fondamentali, sono già disponibili. Il quadro normativo delineato dal binomio RGPD–*AI Act* conferma la possibilità di un approccio integrato, capace di tutelare le libertà individuali e, al contempo, promuovere un'innovazione responsabile. Molto dipenderà dalle modalità concrete di utilizzo, che rifletteranno anche il grado di consapevolezza, da parte delle cosiddette “fabbriche dell'IA”⁹⁴, del necessario passaggio, evidenziato anche dall'*EDPB*, da un'impostazione massimalista in materia di protezione dei dati a un più maturo e realistico bilanciamento tra interessi collettivi e diritti individuali.

⁹² V. AA.VV., *A Survey of Machine Unlearning*, consultabile al sito <https://arxiv.org/abs/2209.02299>.

⁹³ Per un maggiore approfondimento delle tecniche per la tutela della privacy nell'addestramento e diffusione degli LLM si veda AA.VV., *What Does it Mean for a Language Model to Preserve Privacy?*, in *FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022.

⁹⁴ Commissione europea, *Fabbriche di IA*, consultabile al sito <https://digital-strategy.ec.europa.eu/it/policies/ai-factories>.



DARIA LUISA PETRUCCI*

LA TUTELA DEL DIRITTO DELL'INDIVIDUO A UN AMBIENTE SALUBRE NELL'ERA DELL'IA: SFIDE E PROSPETTIVE

SOMMARIO: 1. Introduzione: l'avvento dell'IA nell'attuale contesto di crisi climatica. – 2. Dall'interesse collettivo al diritto soggettivo: il diritto umano a un ambiente salubre e le sue implicazioni giuridiche. – 3. Le Potenzialità dell'IA nella lotta al cambiamento climatico: la *Twin Transition*. – 4. I rischi ambientali legati allo sviluppo dell'IA. – 5. La tutela ambientale e lo sviluppo sostenibile nei nuovi strumenti di regolazione dell'intelligenza artificiale. – 6. L'IA nella transizione verso l'economia circolare. – 7. Conclusioni: verso un equilibrio sostenibile.

1. *Introduzione: l'avvento dell'IA nell'attuale contesto di crisi climatica*

L'intelligenza artificiale sta rivoluzionando il modo in cui l'umanità si approccia alla realtà, incidendo non solo nel campo dell'innovazione e della ricerca, ma anche su numerosi aspetti della vita quotidiana quali la *privacy*, il lavoro, la salute.

Le nuove tecnologie che impiegano sistemi di intelligenza artificiale (d'ora in poi "IA") consentono di elaborare enormi quantità di dati e di svolgere operazioni infinitamente complesse con un elevatissimo grado di accuratezza e in tempi *record*. Per questo motivo, vengono impiegate in un numero sempre crescente di settori, tra i quali figura quello ecologico.

Nel contesto delle sfide contemporanee, quella ambientale è la più imminente: elaborare strategie efficaci e a lungo termine per contrastare il cambiamento climatico e assicurare la conservazione delle risorse e della biodiversità è una necessità alla quale si deve provvedere quanto prima. In termini di tutela ambientale e di sostenibilità molto si discute dell'impiego dell'IA in ambito ecologico e delle potenzialità di questa nuova tecnologia nella lotta al cambiamento climatico e al depauperamento delle risorse. Si parla, in proposito, di *twin transition*¹, espressione con la quale si fa riferimento alla possibilità che l'IA si faccia carico delle problematiche ambientali, indirizzando in modo sostenibile la rivoluzione tecnologica e digitale. Alcune caratteristiche proprie di questi sistemi, quali l'apprendimento,

* Dottoranda in Diritto dell'Unione europea presso l'Università degli Studi di Napoli "Parthenope".

¹ Nel 2020-2021 la Presidente della Commissione europea Ursula von der Leyen ha individuato nelle "transizioni gemelle" il futuro dell'Unione europea.

l'automazione e le straordinarie capacità computazionali, alimentano attualmente il dibattito sulla possibilità che queste nuove tecnologie *smart* possano diventare utili strumenti per la protezione dell'ambiente e la conservazione delle risorse naturali². In quest'ottica, i possibili punti di integrazione tra la rivoluzione digitale dell'IA e la necessità, sempre più impellente, di elaborare nuove strategie di sostenibilità ambientale a lungo termine, rappresentano il fulcro della ricerca scientifica dei prossimi decenni.

Tuttavia, non mancano preoccupazioni circa i potenziali rischi che lo sviluppo delle tecnologie che impiegano sistemi di IA potrebbe comportare per l'ambiente e, di conseguenza, per il diritto del singolo individuo a vivere in un ambiente salubre. Si parla, in proposito, di *hidden ecological cost*, cioè del costo ecologico associato alla nuova rivoluzione digitale, che rimane in gran parte inosservato e non contabilizzato³. Data la straordinaria portata dell'IA, l'analisi del suo impatto ecologico a lungo termine rischia di passare in second'ordine. È invece fondamentale, nell'attuale contesto globale, riconoscere e approfondire il "prezzo" di queste tecnologie in termini ambientali, proprio al fine di assicurarne il miglior impiego possibile e far sì che non si ripeta quanto già accaduto con la rivoluzione industriale, quando lo sviluppo umano in un settore, per l'appunto quello dell'industria, comportò una perdita netta di benessere per la collettività nel settore ambientale.

Il presente contributo, dunque, si propone di analizzare le intersezioni tra l'IA e la tutela ambientale, al fine di indagare se questa possa effettivamente rappresentare un'opportunità per affrontare le minacce poste dalla crisi climatica o se, al contrario, rappresenti un rischio eccessivo per la tutela degli ecosistemi e per il diritto individuale a un ambiente salubre (par. 2). A partire dalle applicazioni pratiche dell'IA (par. 3 e par. 4), si condurrà un'analisi comparativa degli strumenti giuridici e normativi in materia (par. 5), esplorando i limiti e le potenzialità degli stessi (par. 6) e evidenziando possibili sinergie tra i quadri regolatori europei ed internazionali (par. 7).

2. Dall'interesse collettivo al diritto soggettivo: il diritto umano a un ambiente salubre e le sue implicazioni giuridiche

Da quando l'Assemblea Generale delle Nazioni Unite, con la risoluzione n. 76/300 del 28 luglio 2022⁴, ha affermato l'esistenza di un vero e proprio diritto a un ambiente salubre, la prospettiva giuridica della tutela ambientale ha mutato forma. Dopo oltre quarant'anni di evoluzione normativa e giurisprudenziale, che hanno progressivamente delineato la correlazione tra diritti umani e ambiente, la risoluzione ha rappresentato un passaggio simbolico e politico importante nell'affermazione dell'esistenza di un diritto soggettivo a un

² W. D'AVANZO, *Le applicazioni dell'intelligenza artificiale a tutela dell'ambiente*, in *Diritto e giurisprudenza agraria, alimentare e dell'ambiente - DGA* (on-line), n. 2, 2019, p. 1 ss.; F. CAMISA, *Ambiente e tecnologia: l'interconnessione tra le 'transizioni gemelle'*, in *Federalismi.it*, 2024, p. 57; M. OROFINO, *Le due transizioni, digitale e verde, nel cd. pacchetto digitale europeo*, in *Astrid Rassegna*, n. 10, 2023, pp. 1-11.

³ A. ZHUK, *Artificial Intelligence Impact on the Environment: Hidden Ecological Costs and Ethical-Legal Issues*, in *Journal of Digital Technologies and Law*, 2023, p. 932.

⁴ GENERAL ASSEMBLY, *Resolution 76/300, The human rights to a clean, healthy and sustainable environment*, UN Doc. A/RES/76/300, 1 August 2022.

ambiente salubre⁵. L'ambiente, per molto tempo riconosciuto e tutelato come bene pubblico collettivo, diviene l'oggetto di un diritto soggettivo dell'individuo.

Il diritto a un ambiente sano, non espressamente previsto tra quelli tutelati dai maggiori strumenti di protezione dei diritti umani, come la Dichiarazione universale dei diritti dell'uomo o le grandi convenzioni internazionali come la CEDU e la Carta di Nizza, ha avuto per lungo tempo la dignità di principio generale e ha faticato a imporsi come posizione giuridica soggettiva.

Nell'attuale contesto di crisi ecologica, però, è intervenuta la risoluzione dell'Assemblea Generale che, seppur non vincolante, ha segnato un cambiamento di traiettoria nel modo in cui ci si relaziona alla tematica ambientale: le attività altamente inquinanti violano non più "solo" interessi collettivi ambientali, ma anche diritti umani individuali.

Sebbene non esista ancora una fonte universale vincolante che codifichi il diritto a un ambiente sano come diritto umano soggettivo, la risoluzione n. 76/300 è espressione di un convincimento generale degli Stati che si sta progressivamente consolidando, essendo stata sostenuta da 161 Stati dell'ONU. La prassi internazionale passa per la giurisprudenza delle grandi corti regionali, come la Corte Interamericana dei Diritti Umani, che ha riconosciuto espressamente il diritto a un ambiente sano come diritto umano autonomo, valido anche *erga omnes*, con l'opinione consultiva OC-23/17 del 2017⁶. Un ruolo cruciale in questo progressivo percorso di definizione di un diritto umano soggettivo, distinto dalla generica tutela dell'ambiente, è stato svolto dalla giurisprudenza della Corte europea dei diritti dell'uomo, nonostante la CEDU non preveda espressamente tale diritto. A ben vedere, nella Convenzione europea dei diritti dell'uomo, come è logico, non vi è alcun riferimento all'ambiente in generale⁷. Tuttavia, la Corte EDU, a partire dai casi *López Ostra v. Spagna* e *Tatar v. Romania* relativi all'inquinamento e alla protezione ambientale, ha tentato di ovviare a questo *vulnus* normativo interpretando l'art. 8 CEDU (diritto alla vita privata e familiare) in modo da includervi la protezione contro gravi danni ambientali (inquinamento, rumore, rifiuti tossici)⁸.

Inoltre, il diritto umano a un ambiente pulito, sano e sostenibile, è stato incorporato in vari accordi regionali sui diritti umani, tra cui la Carta africana dei diritti dell'uomo e dei popoli, il Protocollo addizionale alla Convenzione americana dei diritti umani nel campo dei diritti economici, sociali e culturali, la Carta araba dei diritti dell'uomo e la Dichiarazione dei

⁵ A livello sovranazionale si è molto discusso dell'esistenza di un diritto soggettivo a un ambiente salubre. Sulla scia della risoluzione n. 48/13 del Consiglio per i Diritti Umani dell'ONU del 2021 (UN Doc. A/HRC/RES/48/13) con la quale è stata riconosciuta formalmente l'esistenza di un diritto umano a un ambiente pulito, sano e sostenibile, la risoluzione n. 76/300 ha rappresentato un punto di svolta nel procedimento di affermazione del diritto stesso, rilevando che «the right to a clean, healthy and sustainable environment is related to other rights and existing international law».

⁶ Corte Interamericana dei Diritti Umani, Opinione Consultiva OC-23/17, 15 novembre 2017.

⁷ Al momento dell'approvazione della Convenzione, nel 1948, non si era ancora sviluppata una coscienza collettiva rispetto al problema ambientale.

⁸ P. MASCARO, *Il diritto a un ambiente salubre nella giurisprudenza della Corte europea dei diritti dell'uomo*, in *Diritto e giurisprudenza agraria, alimentare e dell'ambiente - DGA* (on-line), n. 4, 2023, p.1. Per la giurisprudenza della Corte EDU si v. Corte EDU, sent. 9 dicembre 1994, ric. n. 16798/90, *López Ostra c. Spagna*, § 51; Corte EDU, sent. 19 febbraio 1998, ric. n. 14967/89, *Guerra e altri c. Italia*, §§ 58-60; Corte EDU (GC), sent. 30 novembre 2004, ric. n. 48939/99, *Öneryıldız c. Turchia*, § 87; Corte EDU, sent. 27 gennaio 2009, ric. n. 67021/01, *Tătar c. Romania*, § 107; Corte EDU, sent. 24 gennaio 2019, ric. nn. 54414/13 e 54264/15, *Cordella e altri c. Italia*, § 51.

diritti umani adottata dall'ASEAN⁹. Sul piano domestico, è oggi riconosciuto dalle costituzioni di numerosi Stati¹⁰.

Questo mutamento di paradigma da interesse collettivo a diritto umano soggettivo apre la strada a importanti riflessioni sulla giustizia ambientale, sulla responsabilità delle imprese e sull'effettività delle tutele nell'era della crisi ecologica globale. Nella risoluzione n. 76/300, infatti, l'AG non solo riafferma l'obbligo degli Stati di proteggere e realizzare i diritti umani, riprendendo i Principi quadro sui diritti umani e l'ambiente del 2018¹¹, ma ribadisce anche la responsabilità di tutte le imprese di rispettarli¹², richiamando quanto sancito dai Principi Guida dell'ONU su imprese e diritti umani del 2011¹³ nonché dal rapporto *Mapping* del 2013¹⁴.

Sugli Stati, dunque, grava non solo il dovere di legiferare per stabilire, attuare e mantenere quadri normativi e istituzionali che consentano ai singoli il pieno godimento di un ambiente sano (Principio 11 - Principi quadro del 2018)¹⁵, ma anche l'obbligo di garantire l'effettiva attuazione degli *standard* ambientali nei confronti di attori pubblici e privati (Principio 12 - Principi quadro del 2018), il cd. obbligo di protezione¹⁶. A quanto si deduce dal commentario di quest'ultimo principio, gli Stati devono monitorare e far rispettare efficacemente gli *standard* ambientali adottati prevenendo, investigando, punendo e rimediando alle violazioni dei medesimi compiute dagli attori privati e dalle autorità governative. In particolare, gli Stati devono regolamentare le attività delle imprese per proteggere gli individui dalle violazioni dei diritti umani derivanti da danni ambientali e fornire l'accesso ai rimedi giurisdizionali¹⁷.

Le imprese, a loro volta, sono tenute da una parte a rispettare tutte le normative in vigore in materia ambientale e ad attuare processi di *due diligence* volti a prevenire o a mitigare l'impatto ecologico delle proprie attività, dall'altra sono chiamate ad assumere chiari impegni

⁹ Si v. art. 24 della Carta africana dei diritti dell'uomo e dei popoli; art. 11 Protocollo addizionale alla Convenzione americana dei diritti umani nel campo dei diritti economici, sociali e culturali; art. 38 Carta araba dei diritti dell'uomo; art. 28.f. Dichiarazione dei diritti umani dell'ASEAN.

¹⁰ Stando a quanto riportato dal sito *web* dell'UNEP, oltre un centinaio di Stati hanno incorporato il diritto a un ambiente sano nei propri testi costituzionali.

¹¹ SPECIAL RAPPORTEUR ON THE ISSUE OF HUMAN RIGHTS OBLIGATIONS RELATING TO THE ENJOYMENT OF A SAFE, CLEAN, HEALTHY AND SUSTAINABLE ENVIRONMENT, *Report of Special Rapporteur on the issue of human rights obligations relating to the enjoyment of a safe, clean, healthy and sustainable environment*, UN Doc. A/HRC/40/55, 24, January 2018. I Principi, elaborati nel 2018 dal Relatore Speciale John Knox sistematizzano gli obblighi degli Stati in materia di diritti umani e ambiente.

¹² GENERAL ASSEMBLY, *Resolution 76/300*, cit., considerando 16.

¹³ SPECIAL REPRESENTATIVE OF THE SECRETARY GENERAL ON THE ISSUE OF HUMAN RIGHTS AND TRANSNATIONAL CORPORATIONS AND OTHER BUSINESS ENTERPRISES, *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*, o UNGPs, UN Doc. A/HRC/17/31, 21 March 2011. Gli UNGPs sanciscono l'obbligo degli Stati di proteggere gli individui dalle violazioni dei diritti umani poste in essere da terzi, incluse le imprese.

¹⁴ INDEPENDENT EXPERT ON THE ISSUE OF HUMAN RIGHTS OBLIGATIONS RELATING TO THE ENJOYMENT OF A SAFE, CLEAN, HEALTHY AND SUSTAINABLE ENVIRONMENT, JONH H. KNOX, *Mapping Report*, UN Doc. A/HRC/RES/25/53, 30 December 2013, d'ora in poi rapporto *Mapping* del 2013. Nella sezione dedicata agli obblighi sostanziali in materia di diritti umani e ambiente, il rapporto analizza anche gli obblighi degli Stati volti a regolamentare le attività del settore privato al fine di proteggere gli individui dai danni ambientali associati a tali attività.

¹⁵ V. Principi quadro del 2018, cit. *supra*, nota 11.

¹⁶ *Ibidem*.

¹⁷ Per un approfondimento sul punto v. V. GRADO, *Il diritto umano universale a un ambiente sano: recenti (e futuri) sviluppi*, in *Ordine internazionale e diritti umani*, 2023, p. 229 ss.

di *policy* anche prescindendo dalle normative nazionali di settore¹⁸. Infatti, la responsabilità delle imprese per violazioni dei diritti umani derivanti da danni ecologici esiste “indipendentemente” dalla capacità e/o volontà degli Stati di adempiere ai propri obblighi in materia di ambiente, come precisato dal commentario del Principio 11 degli UNGPs del 2011¹⁹.

Un ulteriore sviluppo significativo è rappresentato dal recente Parere consultivo della Corte internazionale di giustizia del 23 luglio 2025²⁰, emesso nell'ambito della richiesta sul contenuto e sulla portata degli obblighi degli Stati in relazione al cambiamento climatico.

La Corte ha affermato in modo esplicito che «the human right to a clean, healthy and sustainable environment is essential for the enjoyment of other human rights», chiarendo così, a livello universale, che un ambiente pulito, sano e sostenibile costituisce una precondizione necessaria per l'effettivo godimento di altri diritti umani fondamentali, tra cui il diritto alla vita e alla salute.

Il parere completa la traiettoria evolutiva già tracciata dalla Risoluzione dell'Assemblea Generale ONU 76/300 del 28 luglio 2022: mentre quest'ultima ha natura politica e riconosce l'ambiente sano come diritto umano in termini di indirizzo, la Corte internazionale di giustizia ne conferma la rilevanza giuridica, riconoscendo che l'obbligo degli Stati di prevenire danni ambientali significativi discende non solo dagli accordi internazionali, ma anche dal diritto internazionale consuetudinario e dal nucleo essenziale dei diritti umani.

La Corte precisa, inoltre, che la responsabilità internazionale è conseguenza non solo della violazione degli obblighi climatici, ma anche della mera inazione: la mancata adozione di misure dovute, quando sussista un obbligo di prevenzione, può infatti integrare un illecito internazionale. Tali affermazioni rafforzano la qualificazione dell'ambiente salubre come diritto fondamentale dotato di un contenuto autonomo ed effettivo, la cui protezione richiede comportamenti positivi da parte degli Stati.

3. Le Potenzialità dell'IA nella lotta al cambiamento climatico: la Twin Transition

Secondo alcuni esperti di transizione ecologica²¹, attraverso la riduzione degli sprechi, l'ottimizzazione delle risorse e il monitoraggio ambientale, l'IA potrebbe disegnare una nuova traiettoria verso la sostenibilità, implementando al tempo stesso due transizioni: quella tecnologica e quella ambientale. Sono infatti molteplici le potenzialità di questa tecnologia in ambito ecologico²².

¹⁸ Tra i documenti già citati, fanno specifico riferimento alla responsabilità delle imprese di rispettare i diritti umani, oltre agli UNGPs e alla recente ris. 76/300 dell'AG, anche il rapporto *Mapping* del 2013 e i Principi quadro del 2018.

¹⁹ V. UNGPs del 2011, cit. *supra*, nota 17. Sul punto v. anche V. GRADO, *Il diritto umano universale a un ambiente sano*, p. 234.

²⁰ Corte int. giust., Parere consultivo 23 luglio 2025, *Obblighi degli Stati in relazione al cambiamento climatico*.

²¹ M. R. WADE, K. TRANTOPOULOS, *Twin Transformation: How to Stop Treating AI and Sustainability as Separate Challenges*, 2025.

²² Sugli aspetti positivi della trasformazione digitale e ambientale W. D'AVANZO, *Algoritmi, smart cities, metaversi. Etica ambientale ed etica digitale per lo sviluppo sostenibile*, in *Diritto e giurisprudenza agraria, alimentare e dell'ambiente - DGA* (on-line), n. 2, 2024, p. 1 ss.; W. D'AVANZO, *Le applicazioni dell'intelligenza artificiale* cit., p. 1 ss.; F. CAMISA, *Ambiente e tecnologia*, cit., p. 56 ss.

Un primo aspetto riguarda il monitoraggio dei livelli di inquinamento e di emissioni²³. L'IA, infatti, può contribuire a ridurre l'impatto ambientale delle attività umane attraverso strumenti di analisi avanzata e di previsione, nonché attraverso l'elaborazione di strategie di ottimizzazione delle risorse e analisi di *environmental risk management*²⁴. In tale settore, avanguardistico risulta l'impiego delle cosiddette *smart grids* o reti intelligenti, sistemi di distribuzione dell'energia elettrica che utilizzano *software* di IA per ottimizzare la produzione, la distribuzione e il consumo di energia e migliorare l'efficienza, l'affidabilità e la sostenibilità della rete elettrica²⁵.

Altro impiego dei sistemi di IA riguarda il monitoraggio e la prevenzione dei disastri ambientali: i *software* di IA potrebbero essere impiegati per rilevare tempestivamente anomalie di vario genere tramite l'analisi delle immagini satellitari, consentendo così di attuare interventi immediati²⁶. Si pensi al fenomeno, piuttosto frequente, dell'*oil spilling*, rispetto al quale un tempestivo intervento per frenare le fuoriuscite di petrolio è fondamentale.

Altrettanto rilevante per la tutela ambientale è l'utilizzo dell'IA nel settore della conservazione della biodiversità²⁷. Attualmente esistono già progetti in Italia²⁸ che utilizzano dispositivi acustici basati sull'IA per monitorare oasi protette dal WWF e specie a rischio, rilevando eventuali attività illegali in tempo reale²⁹.

E ancora, nel campo della sicurezza alimentare, alcuni recenti studi condotti da ENEA³⁰ e dall'Università degli Studi di Bari³¹ hanno impiegato tecniche di *machine learning* per analizzare il microbioma del suolo, identificando indicatori di salute del terreno e contribuendo alla sicurezza alimentare³².

In fine, la gestione dei rifiuti e la raccolta differenziata potrebbero certamente beneficiare dell'utilizzo delle tecnologie di IA. In particolar modo, il riciclo e la riprogettazione dei rifiuti in nuovi prodotti aiuterebbero a promuovere il buon funzionamento dell'economia circolare³³.

²³ M. SINGH, S. SINGH, *Role of Artificial Intelligence for Environment Protection: An Analysis*, in *Law Journal of GNDU*, 2023, p. 172; A. P. VICTORIO, E. CELESTRE, A. QUINTAVALLA, *Greening AI? The New Principle of Sustainable Digital Products and Services in the EU*, in *CMLR*, 2024, p. 1019 ss.

²⁴ W. D'AVANZO, *Environmental risk management e intelligenza artificiale simbiotica. Prospettive di sviluppo e criticità*, in *Diritto e giurisprudenza agraria, alimentare e dell'ambiente - DGA* (on-line), n. 2, 2025, p. 1 ss.; M. C. GAETA, *Intelligenza artificiale sostenibile e tutela dei green Rights*, in *European Journal of Privacy Law & Technologies – EJPLT* (on-line), 2024, p.148.

²⁵ Sul punto T. LYNN, P. ROSATI, D. KREPS, K. CONBOY, *Digital Sustainability – Leveraging Digital Tech to Combat Climate Change*, *Palgrave Studies*, in *Digital Business and Enabling Technologies*, 2024.

²⁶ K. OTSU, J. MASO, *Digital Twins for Research and Innovation in Support of the European Green Deal Data Space: A Systematic Review*, in *Remote Sensing*, 2024, p. 4 ss. Questo studio descrive strumenti digitali avanzati per il monitoraggio dei disastri ambientali tramite satelliti. In particolare, presenta *Destination Earth*, un'iniziativa della Commissione europea per sviluppare un gemello digitale della Terra ad alta precisione, capace di monitorare e prevedere l'evoluzione dei fenomeni naturali e delle attività umane sul pianeta.

²⁷ Sul punto A. PACHOT, A. PATISSIER, *Towards Sustainable Artificial Intelligence: An Overview of Environmental Protection Uses and Issues*, in *Green and Low-Carbon Economy*, 2022.

²⁸ Maggiori informazioni sul progetto *Nature Guardian*, frutto della collaborazione tra WWF Italia, Huawei e Rainforest Connection, sono reperibili sul sito ufficiale del WWF o sul sito di Rainforest Connection.

²⁹ *Ibidem*.

³⁰ Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo sostenibile.

³¹ G. CARRERA, R. BARBATO, M. LIBERATI, M. ROMANO, *Machine Learning for Soil Microbiome Analysis to Assess Climate Change Effects*, in *Machine Learning and Knowledge Extraction*, vol. 6, n. 4, 2024, pp. 736–749.

³² T. LYNN et al., *Digital Sustainability*, cit.

³³ Sul punto A. PACHOT, A. PATISSIER, *Towards Sustainable Artificial Intelligence*, cit.

4. I rischi ambientali legati allo sviluppo dell'IA

Se da un lato è indiscutibilmente vero che l'IA potrebbe rappresentare un *turning point* nella lotta al depauperamento delle risorse, contribuendo ad individuare strategie a lungo termine per la risoluzione delle sempre più imminenti sfide ambientali, d'altro molto si discute dei potenziali rischi che lo sviluppo di queste tecnologie avanguardistiche potrebbero determinare rispetto alla tutela ambientale, con inevitabili ripercussioni sul diritto soggettivo a un ambiente salubre³⁴. Infatti, l'esecuzione e l'addestramento di modelli di IA richiedono l'impiego di smisurate quantità di energia e di risorse idriche, rischiando di generare un danno ambientale non riassorbibile che comprometterebbe, nel lungo periodo, la qualità di vita della generalità degli individui in maniera irrimediabile.³⁵

In primo luogo, rileva la questione dei *data centres*, enormi complessi industriali nei quali vengono sviluppate e *trainate* le tecnologie che impiegano IA, che contribuiscono in modo significativo al consumo energetico e che sono spesso alimentati da fonti di energia non rinnovabili³⁶. Queste strutture richiedono una notevole quantità di elettricità per alimentare i *server* e, allo stesso tempo, anche un dispendioso impiego di risorse idriche per i sistemi di raffreddamento delle apparecchiature di rete. Si stima che un utilizzo continuo dei *server* possa richiedere oltre 400 milioni di litri d'acqua all'anno³⁷.

Anche la fase dell'addestramento dei sistemi di IA richiede enormi quantità di energia per processare tutti i dati che vengono impiegati nel *training*. Infatti, i calcoli computazionali alla base di queste tecnologie e lo sviluppo di complessi algoritmi di funzionamento automatizzato richiedono l'elaborazione di enormi quantità di dati, processi energivori che, entro il 2026, potrebbero richiedere un consumo totale di elettricità a livello globale superiore ai 1.000 TWh, equivalendo approssimativamente al consumo elettrico del Giappone³⁸. Il consumo di energia durante l'addestramento può variare da alcune centinaia di chilowattora (kWh) a diverse migliaia di kWh, a seconda delle dimensioni e della complessità del modello, del *set* di dati e dell'infrastruttura *hardware* utilizzata³⁹.

³⁴ Sulle ripercussioni della trasformazione digitale P. DE PASQUALE, *Sostenibilità e trasformazione digitale: paradigmi a confronto nella disciplina dell'Unione europea*, in DUE, 2022, p. 67 ss.; G. PIRINA, *I costi umani e ambientali del digitale*, in *il Mulino*, n. 3, 2022, pp. 57-64; M. TOMASI, *Intelligenza artificiale, sostenibilità e responsabilità intergenerazionali: nuove sfide per il costituzionalismo?*, in *Rivista AIC*, n. 4/2024, p. 49.

³⁵ Sulle esternalità ambientali negative della trasformazione digitale si veda C. DI FRANCESCO MAESA, *Economia circolare e IA: a che condizioni è una sfida possibile per l'UE?*, in *Quaderni AISDUE - Rivista quadrimestrale* ISSN 2975-2698 – *Anticipazione fascicolo*, Convegno Forum AES del 5 e 6 dicembre 2024, p. 10 ss.

³⁶ P. VICTORIO, E. CELESTRE, A. QUINTAVALLA, *Greening AI?*, cit. p. 1024; A. ZHUK, *Artificial Intelligence Impact on the Environment*, cit., p. 938. Anche SERBAN et al., nel suo studio *Artificial intelligence for smart renewable energy sector in Europe - smart energy infrastructures for next generation smart cities*, 2020, sottolinea l'importanza di passare a fonti di energia rinnovabili.

³⁷ P. LI et al., *Making AI Less 'Thirsty': Uncovering and Addressing the Secret Water Footprint of AI Models*, in *arXiv*, 2023; A. ZHUK, *Artificial Intelligence Impact on the Environment*, cit., p. 938.

³⁸ Sul consumo energetico dei *data center* si veda *report* dell'INTERNATIONAL ENERGY AGENCY (IEA), *Electricity 2024: Analysis and Forecast to 2026*, del 2024. Stando a quanto riportato dal *report*, nel 2023 i *data center* hanno consumato circa 460 TWh a livello globale. Questo valore è destinato ad aumentare fino a 620-1.050 TWh nel 2026, a seconda degli scenari considerati. Una porzione significativa di questa crescita è attribuita, in particolare, all'addestramento e all'uso di modelli di IA generativa, come i Large Language Models (LLMs).

³⁹ A. ZHUK, *Artificial Intelligence Impact on the Environment*, cit., p. 936. Sul punto anche R. SCHWARTZ, J. DODGE, N. A. SMITH, O. ETZIONI, *Green AI*, in *arXiv*, 2020.

Oltre al problema dell'impiego delle risorse, c'è da esaminare anche la contropartita rappresentata dalle emissioni di gas nocivi che ne derivano (*digital pollution*). Basti pensare che l'addestramento di un singolo modello linguistico di IA può arrivare ad emettere tanta anidride carbonica quanta ne viene emessa nel corso della vita di cinque automobili⁴⁰ e che, più in generale, lo sviluppo di un singolo modello di IA può comportare l'emissione di circa 300 tonnellate di CO₂⁴¹.

Un altro aspetto da prendere in considerazione riguarda la fabbricazione di componenti necessari per il funzionamento di sistemi di IA, come le GPU⁴², che implica l'estrazione di materie prime e la generazione esponenziale di rifiuti elettronici⁴³. La rapida evoluzione dei dispositivi utilizzati per produrre tecnologia di IA, infatti, comporta la rapida obsolescenza di questi *hardware*, che presentano cicli di vita sempre più brevi. Ciò genera una crescente quantità di *e-waste*, spesso non adeguatamente riciclati, con catastrofici impatti sull'ambiente⁴⁴. Il rapporto *Global E-waste Monitor 2020* indica che la produzione di rifiuti elettronici ha raggiunto la cifra record di 53,6 milioni di tonnellate, con solo il 17,4% raccolto e riciclato ufficialmente⁴⁵.

Nonostante il moltiplicarsi degli studi, i rischi ambientali associati all'impiego delle nuove tecnologie di IA non solo permangono, ma aumentano. In risposta all'impennata della domanda, le *big tech* ne stanno incrementando la produzione e, in conseguenza, aumenterà anche il consumo energetico necessario per alimentare le infrastrutture di elaborazione. Inoltre, le numerose applicazioni di sistemi di IA continuano a progredire e a diventare sempre più complesse, comportando un aumento del numero e delle dimensioni dei *data center* e un incremento della potenza computazionale e di calcolo richiesta, con spaventose ripercussioni sul piano energetico⁴⁶.

Non bisogna, inoltre, sottovalutare i rischi connessi a un cattivo impiego delle stesse. L'uso inappropriato o non regolamentato di queste tecnologie, sempre più alla portata di mercato, potrebbe vanificare i richiamati benefici connessi al monitoraggio ambientale e agli sforzi di conservazione delle risorse e della biodiversità, sconvolgendo il naturale funzionamento degli ecosistemi naturali. Basti pensare che, secondo gli esperti, i droni per il monitoraggio ambientale e i veicoli autonomi utilizzati per l'esplorazione delle risorse potrebbero potenzialmente disturbare gli *habitat* della fauna selvatica, interferendo con i modelli di migrazione e aggravando gli squilibri dell'ecosistema. Si parla in proposito di *unintended consequences of AI on biodiversity and ecosystems*⁴⁷.

⁴⁰ *Ibidem*.

⁴¹ E. STRUBELL, A. GANESH, A. MCCALLUM, *Energy and Policy Considerations for Deep Learning in NLP*, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 3645–3650; R. SCHWARTZ, J. DODGE, N. A. SMITH, O. ETZIONI, *Green AI*, in *arXiv*, 2020.

⁴² Le GPU (*Graphics Processing Unit*) sono unità di elaborazione grafica, ovvero processori specializzati per elaborare immagini (come quelle dei *videogames* o delle animazioni 3D), che oggi vengono ampiamente utilizzate anche al di fuori dell'ambito grafico, fungendo da “motore” computazionale che rende possibile l'addestramento e il funzionamento dell'IA moderna.

⁴³ DE PASQUALE, *Sostenibilità e trasformazione digitale*, cit., p. 75.

⁴⁴ P. WANG et al., *E-waste Challenges of Generative Artificial Intelligence*, in *Journal of Digital Technologies and Law*, 2024, p. 1 ss.

⁴⁵ V. FORTI, C.P. BALDÉ, R. KUEHR, G. BEL, *The Global E-waste Monitor 2020: Quantities, flows and the circular economy potential*, United Nations Institute for Training and Research (UNITAR) & International Telecommunication Union (ITU), 2020, p. 4.

⁴⁶ P. DE PASQUALE, *Sostenibilità e trasformazione digitale*, cit., p. 85; A. ZHUK, *Artificial Intelligence Impact on the Environment*, cit., p. 935.

⁴⁷ *Ibidem*.

5. La tutela ambientale e lo sviluppo sostenibile nei nuovi strumenti di regolazione dell'intelligenza artificiale

La tutela dell'ambiente rientra tra gli obiettivi che devono orientare l'azione dell'Unione europea ai sensi dell'art. 3.3 del TUE. L'art. 37 della Carta dei diritti fondamentali dell'Unione e l'art. 11 del TFUE sanciscono, inoltre, il principio di integrazione delle esigenze ambientali in tutte le politiche e azioni dell'Unione.

Accanto a questo nucleo di disposizioni di diritto primario, il quadro normativo dell'Unione si è recentemente arricchito di atti di diritto derivato volti a rafforzare l'integrazione degli obiettivi ambientali nelle attività economiche, quali la direttiva sulla sostenibilità⁴⁸ e la direttiva sulla responsabilità delle imprese⁴⁹.

Parallelamente, la direzione politica dell'Unione deve essere orientata al rispetto di iniziative *Green oriented*, tra cui il *Green Deal*⁵⁰ e la Dichiarazione europea sui diritti e i principi digitali, il cui Capitolo VI, rubricato *Sostenibilità*, è espressamente dedicato all'impatto ambientale delle tecnologie digitali.⁵¹ Tali strumenti di *soft law*, seppur non vincolanti, contribuiscono a delineare la cornice entro cui si collocano le recenti iniziative legislative in materia di sviluppo sostenibile e di responsabilità delle imprese.⁵²

Anche l'*AI ACT*⁵³, che rappresenta il primo tentativo organico di regolamentazione dell'intelligenza artificiale, si impegna ad assicurare che lo sviluppo e l'impiego dei sistemi di IA siano sostenibili da un punto di vista ambientale⁵⁴. Tra gli obiettivi del nuovo regolamento, infatti, figura quello di garantire un elevato livello di protezione dei diritti fondamentali riconosciuti e tutelati dal diritto dell'Unione, tra cui la democrazia, lo Stato di diritto e la tutela dell'ambiente⁵⁵. Il riferimento alla necessità che l'implementazione di queste nuove tecnologie rispetti l'ambiente è frequente all'interno del regolamento, tanto nei considerando⁵⁶ quanto nelle norme⁵⁷, ma non vengono specificate le modalità. Sul punto, il reg. (UE) 2024/1689 si limita a generici richiami.⁵⁸ D'altronde, l'*AI ACT* si occupa prevalentemente di regolare la circolazione dell'IA all'interno del mercato e non introduce alcun coordinamento con il *corpus* normativo dell'Unione a tutela dell'ambiente.

Al considerando 27, per esempio, si evidenzia la necessità che i sistemi di IA siano sviluppati e utilizzati, tra l'altro, «in modo sostenibile e rispettoso dell'ambiente [...]». Il

⁴⁸ Direttiva (UE) 2022/2464 del Parlamento europeo e del Consiglio, del 14 dicembre 2022, che modifica il regolamento (UE) n. 537/2014, la direttiva 2004/109/CE, la direttiva 2006/43/CE e la direttiva 2013/34/UE in materia di rendicontazione societaria di sostenibilità.

⁴⁹ Direttiva (UE) 2024/1760 del Parlamento europeo e del Consiglio, del 5 luglio 2024, relativa al dovere di diligenza delle imprese ai fini della sostenibilità (*Corporate Sustainability Due Diligence Directive* – CSDDD).

⁵⁰ Comunicazione della Commissione COM(2019) 640 final, *Il Green Deal europeo*, 11 dicembre 2019.

⁵¹ P. VICTORIO, E. CELESTRE, A. QUINTAVALLA, *Greening AI?*, cit. p. 1021.

⁵² P. DE PASQUALE, *Sostenibilità e trasformazione digitale*, cit., p. 70.

⁵³ Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828.

⁵⁴ F. CAMISA, *Ambiente e tecnologia*, cit., p. 69.

⁵⁵ M. ALMADA, N. PETTIT, *The EU AI act: a medley of product safety and fundamental rights?*, European University Institute, Robert Schuman Centre Working Paper, 59, 2023, pp. 7-27.

⁵⁶ Reg. (UE) 2024/1689 - considerando 1, 2, 4, 8, 27, 48, 130, 142, 155, 176.

⁵⁷ *Ivi*, art. 1, art. 3, art. 46, art. 59.

⁵⁸ C. DI FRANCESCO MAESA, *Economia circolare*, cit., p. 16 ss.

considerando 48, invece, fa espressamente riferimento al «diritto fondamentale a un livello elevato di protezione dell'ambiente» tra quelli da tenere in considerazione nella valutazione circa la gravità di un ipotetico danno causato da un sistema di IA. Il considerando 142 incoraggia gli Stati a «sostenere e promuovere la ricerca e lo sviluppo di soluzioni di IA a sostegno di risultati vantaggiosi dal punto di vista sociale e ambientale [...] e a conseguire obiettivi in materia di ambiente». Il considerando 155 precisa che i fornitori di sistemi di IA ad alto rischio «dovrebbero [...] essere tenuti [...] a predisporre un sistema per segnalare alle autorità competenti eventuali incidenti gravi», tra cui sono compresi i «gravi danni alle cose o all'ambiente».

Sul piano operativo, la norma che instaura un tenue coordinamento di tipo pratico con le altre disposizioni ambientali dell'Unione è l'articolo 40, che prevede l'intervento della Commissione per promuovere la *standardizzazione* tecnica. Il comma 2 stabilisce che la Commissione debba richiedere agli organismi di *standardizzazione* di fornire risultati relativi «a processi di comunicazione e documentazione intesi a migliorare le prestazioni dei sistemi di IA in termini di risorse, come la riduzione del consumo di energia». In questo senso, l'implementazione della sostenibilità energetica dell'IA è rimessa allo sviluppo di *standard* armonizzati da parte degli organismi competenti. Questa impostazione riflette un modello di regolazione *soft*, basato sull'autodisciplina tecnica attraverso l'implementazione di *standard* condivisi piuttosto che su obblighi immediatamente vincolanti. Tuttavia, questa strategia di «rinvio» genera un duplice rischio: da un lato, i tempi di approvazione degli *standard* da parte degli organismi competenti sono spesso lunghi; dall'altro, tali organismi fanno ampio riferimento a *stakeholder* industriali, con il rischio di affidare alle logiche di mercato la definizione di regole che dovrebbero invece garantire un bilanciamento effettivo tra innovazione tecnologica e sostenibilità ambientale.⁵⁹

Oltre all'articolo 40, un ulteriore obbligo in materia ambientale è contenuto nell'Allegato XI⁶⁰ e riguarda specificamente i modelli di IA per finalità generali. In base a tale disposizione, i fornitori di tali modelli sono tenuti a comunicare, nelle informazioni tecniche, il consumo energetico noto o stimato associato alla fase di addestramento del sistema. In presenza di rischi sistemici, è altresì previsto che si effettui una valutazione dell'efficienza energetica del modello, con possibilità per la Commissione di elaborare metodologie comuni di misurazione al fine di garantire comparabilità e coerenza tra i diversi sviluppatori.⁶¹ Tali forme di *disclosure* si inseriscono in un tentativo di aumentare la trasparenza sull'impronta ecologica delle architetture più potenti e pervasive, in assenza, tuttavia, di veri e propri vincoli sostanziali in termini di riduzione dei consumi o minimizzazione dell'impatto.

Rilevate queste criticità, l'*AI ACT* mostra una debolezza strutturale rispetto ad altri atti normativi dell'Unione che adottano approcci più stringenti e orientati a una piena integrazione tra transizione digitale e transizione verde.⁶² La direttiva CSDDD (*Corporate Sustainability Due Diligence Directive*), per esempio, adotta un approccio prescrittivo, imponendo obblighi stringenti alle imprese in materia ambientale lungo tutta la catena del valore. La direttiva 2022/2464 in materia di rendicontazione societaria di sostenibilità, invece, impone alle imprese l'obbligo di fornire una dettagliata informativa non finanziaria sulle *performance* ambientali, sociali e di *governance*, rafforzando così il principio di trasparenza e il controllo pubblico sulle strategie aziendali in materia di sostenibilità. Entrambe delineano un

⁵⁹ M. TOMASI, *Intelligenza artificiale*, cit., p. 56.

⁶⁰ Reg. (UE) 2024/1689 - Allegato XI, par. 2, lett. e).

⁶¹ *Ivi*, art. 112, par. 6; art. 53, par. 1, lett. a) in combinato disposto con l'Allegato XI.

⁶² C. DI FRANCESCO MAESA, *Economia circolare*, cit., p. 18.

quadro normativo più ambizioso e vincolante rispetto all'*AI ACT* che, sul piano ambientale, resta ancorato a mere indicazioni programmatiche.

Ancora più generica è la Convenzione quadro del Consiglio d'Europa sull'intelligenza artificiale e i diritti umani, la democrazia e lo Stato di diritto, adottata a Strasburgo il 17 maggio 2024. All'interno del Preambolo, si specifica «this Convention is intended to address specific challenges which arise throughout the lifecycle of artificial intelligence systems and encourage the consideration of the wider risks and impacts related to these technologies including, but not limited to, human health and the *environment* [...]». Per quanto concerne la parte normativa, composta da sole 36 norme, l'ambiente figura all'articolo 19, rubricato «Consultazione pubblica» che recita «Ciascuna Parte si adopera per garantire che le questioni importanti sollevate in relazione ai sistemi di intelligenza artificiale siano, [...], debitamente esaminate [...], alla luce delle implicazioni sociali, economiche, giuridiche, etiche, *ambientali* e di altro tipo».

Nonostante l'interesse dimostrato verso il tema ambientale, l'applicabilità della Convenzione è oggi limitata. Al momento solo pochi Stati⁶³ l'hanno firmata e si prevede che non sarà operativa nel breve periodo, non essendo stato ancora depositato alcuno strumento di ratifica⁶⁴. Resta tuttavia auspicabile che la Convenzione venga ratificata e attuata, atteso che la cooperazione internazionale risulta indispensabile per affrontare efficacemente le questioni ambientali derivanti dall'IA. In definitiva, per il momento, il quadro rimane prematuro, e la sua effettiva implementazione potrebbe richiedere tempo e ulteriori passi concreti da parte degli Stati.

6. L'IA nella transizione verso l'economia circolare

Da quanto detto finora, emerge come la trasformazione digitale porti con sé indiscutibili vantaggi e altrettante sfide, che possono essere sommariamente ricondotte a due macrocategorie: da una parte si colloca il tema delle esternalità ambientali, dall'altra quello del depauperamento delle risorse.

Con riguardo all'impoverimento del capitale naturale, molti studi rilevano che l'elevato consumo energetico e idrico che deriva dall'addestramento e dall'impiego di modelli di IA comporta un dispendio di risorse naturali che il Pianeta non può permettersi e che non sarà sostenibile nel tempo. La soluzione prospettata da alcuni studiosi consiste nel riconvertire l'industria produttiva dell'IA in un settore *eco-friendly*, alimentato esclusivamente da fonti di energia rinnovabile. Se si riuscisse a orientare la produzione digitale verso l'impiego di energie pulite, sarebbe possibile continuare a sviluppare nuove tecnologie basate sull'IA senza aggravare il problema della scarsità delle risorse⁶⁵.

⁶³ Hanno firmato il trattato quattordici Stati e l'Unione europea.

⁶⁴ È necessario il raggiungimento di cinque ratifiche (di cui almeno tre provenienti da Stati membri del Consiglio d'Europa) perché entri in vigore.

⁶⁵ A. ZHUK, *Artificial Intelligence Impact on the Environment*, cit., p. 932 ss.; E. ZHANG, D. WU, J. BOMAN, *Carbon-Aware Workload Shifting for Mitigating Environmental Impact of Generative AI Models*, in *Proceedings of IEEE GreenCom 2024*, p. 446 ss.; A. PACHOT, A. PATISSIER, *Towards Sustainable Artificial Intelligence*, cit.; M. A. STEIN, *Artificial Intelligence and Climate Change: Towards an Ethical Legal Framework*, in *International Journal of Environmental Law*, 2023, n. 2, p. 77 ss.; P. LI et al., *Making AI Less 'Thirsty'*, cit.; A. MORAIN, N. ILANGOVAN, C. DELHOM et al., *Artificial Intelligence for Water Consumption Assessment: State of the Art Review*, in *Water Resour Management*, 2024; OECD, *Measuring the Environmental Impacts of AI Compute and Applications*, in *OECD Digital Economy Papers*, n. 341, 2022;

Rispetto alla questione dell'inquinamento ambientale, recenti studi evidenziano come un corretto impiego delle tecnologie di IA possa contribuire, nel lungo periodo, all'elaborazione di nuove politiche economiche in grado di mitigare l'impatto ambientale⁶⁶. Secondo questa prospettiva, dopo un iniziale effetto negativo dell'IA sull'ecosistema, il danno potrebbe progressivamente ridursi grazie a sistemi di IA in grado di ottimizzare i processi produttivi e migliorare l'efficienza energetica, favorendo così un parziale "riassorbimento" del danno⁶⁷. Tuttavia, l'elaborazione di strategie di efficientamento, per quanto utile, non può da sola annullare né compensare integralmente i danni già prodotti. Anzi, l'attuale tendenza mostra un incremento costante delle emissioni associate all'addestramento e all'utilizzo dei modelli di IA e solo scenari ottimistici ipotizzano una loro effettiva riduzione, eventualmente resa possibile dall'uso esclusivo di fonti energetiche rinnovabili.⁶⁸ Inoltre, va sottolineato che l'IA non ha alcun effetto correttivo diretto rispetto all'erosione delle risorse naturali, in particolare quelle idriche ed energetiche. Infine, occorre ricordare che alcuni danni ambientali sono, per loro natura, irreversibili: tra questi, la perdita di biodiversità, lo scioglimento dei ghiacci perenni e l'acidificazione degli oceani⁶⁹.

Ad ogni modo, con riguardo all'impatto ecologico dell'IA, si dovrebbe tendere a un approccio preventivo piuttosto che "curativo". In questo senso viene in rilievo la distinzione, già oggetto di critica da parte della dottrina⁷⁰, tra *sustainable AI* e *AI for sustainability*. In risposta alla massiccia proliferazione delle tecnologie di IA si promuove l'implementazione di una *sustainable AI*, ovvero di un'IA che tenda al rispetto dell'ecosistema, ma non si discute ancora abbastanza di un *AI for sustainability*, cioè di un'IA che, nel suo agire, non si limiti a rispettare gli argini derivanti dalla tutela ambientale, ma che sia direttamente finalizzata alla promozione dello sviluppo sostenibile. La ricerca dovrebbe mirare non solo a rendere sostenibile la trasformazione digitale, ma ancor più a porre l'IA al servizio dell'economia circolare. In tal modo, si realizzerebbe davvero il passaggio dall'industria 4.0 all'industria 5.0, intesa come un modello economico-industriale di integrazione uomo-macchina, in cui l'ambiente non venga compromesso e nel quale si realizzi un effettivo *surplus* di benessere per la collettività nel suo insieme⁷¹.

In questo scenario, l'IA potrebbe rappresentare uno strumento strategico per lo sviluppo dell'economia circolare. L'adozione di sistemi di IA capaci di ottimizzare il riciclo, ridurre gli scarti e monitorare l'intero ciclo di vita dei prodotti permetterebbe di chiudere i cicli produttivi in modo intelligente e responsabile. La combinazione tra IA e modelli circolari

A. COCCA GUARDIA, L. TISCORNIA, A. VELÁZQUEZ, *Assessing the Energy Impact and Carbon Footprint of AI Model Training: A Case Study Using GPU Servers*, in *Proceedings of CLAGTEE 2024*, 2024, E. STRUBELL et al., *Energy and Policy Considerations*, cit., pp. 3645–3650.

⁶⁶ F. CAMISA, *Ambiente e tecnologia*, cit., p. 58.

⁶⁷ K. OTSU, J. MASO, *Digital Twins*, cit., p. 3-4; T. LYNN et al., *Digital Sustainability*, cit.; R. RAMAN et al., *Green and sustainable AI research: an integrated thematic and topic modeling analysis*, in *Journal of Big Data*, 2024; EUROFER, *Industry 4.0 and the Road to Sustainable Steelmaking in Europe - Recasting the Future*, 2024, p. 5-6; A. VERDECCHA, *Reinforcing Sustainability Awareness in the AI Ecosystem: A Strategic View on the AI Value Chain*, in *Sustainability*, n. 5, 2024; M. SINGH, S. SINGH, *Role of Artificial Intelligence*, cit., p. 165.

⁶⁸ P. VICTORIO, E. CELESTRE, A. QUINTAVALLA, *Greening AI?*, cit. p. 1039.

⁶⁹ P. LI et al., *Making AI Less 'Thirsty'*, cit.; E. STRUBELL et al., *Energy and Policy Considerations*, cit., p. 3645–3650.; M. A. STEIN, *Artificial Intelligence and Climate Change*, cit., p. 77 ss.

⁷⁰ M. C. GAETA, *Intelligenza artificiale sostenibile*, cit., p. 148; M. TOMASI, *Intelligenza artificiale*, cit., p. 57; A. VAN WYNSBERGHE, *Sustainable AI: AI for sustainability and the sustainability of AI*, in *AI and Ethics*, 2021, I, p. 214.

⁷¹ K. REJEB; J. G. KEOGH; E. SÜLE, *When Industry 5.0 Meets the Circular Economy: A Systematic Literature Review of Industry 5.0 and CE*, in *Circular Economy and Sustainability*, vol. 6, 2025; P. J. AMIRKHIZI, S. PEDRAMMEHR, S. PAKZAD, *A system-based approach to value prioritization in Industry 5.0*, in *Discover Sustainability*, vol. 6, 2025.

consentirebbe non solo di ridurre l'impronta ecologica dell'industria, ma anche di promuovere nuovi paradigmi economici rigenerativi. Si tratta di una sinergia che potrebbe contribuire concretamente alla costruzione di un ecosistema produttivo, resiliente e orientato alla neutralità climatica.

È chiaro che si tratta di considerazioni acerbe: occorrono ulteriori anni di ricerca e di approfondimento sulla tematica; ma ciò che fin da subito appare chiaro è che la *twin transition* assume oggi un ruolo chiave nell'ambito della ricerca scientifica, rappresentando il punto di partenza per l'implementazione di nuovi obiettivi di prioritaria importanza, diretti a tutelare la salute e al benessere di tutti gli individui, anche di quelli non ancora venuti ad esistenza.

7. Conclusioni: verso un equilibrio sostenibile

In conclusione, se è pacifico che l'impiego dell'IA rappresenti un'opportunità epocale, è altrettanto evidente che una simile innovazione debba essere analizzata e conseguentemente normata per quello che è: uno strumento potente, in grado di innescare un'irreversibile trasformazione della società contemporanea e del Pianeta che la ospita. In questo contesto è fondamentale che l'evoluzione tecnologica proceda all'interno di un perimetro normativo ed etico che garantisca il rispetto dei diritti fondamentali, tra cui il diritto a un ambiente sano. Lo sviluppo umano ha nuovamente condotto a un *trade-off* progresso tecnologico-tutela ecologica e il modo in cui questa tecnologia sarà sfruttata in futuro dipende dal legislatore, la cui sfida principale sarà quella di conciliare la transizione digitale con quella ecologica, promuovendo applicazioni dell'IA che contribuiscano al bene comune e arginando gli impieghi potenzialmente lesivi.

In quest'ottica il quadro europeo, pur dimostrando una crescente attenzione verso la sostenibilità ambientale, come nel caso dell'*AI ACT*, presenta margini di miglioramento.⁷² Come si è detto, il regolamento si limita a raccomandare che lo sviluppo e l'impiego dei sistemi di IA siano sostenibili sotto il profilo ecologico e a fornire indicazioni generali, senza introdurre criteri operativi o vincoli precisi.

In tal senso, al fine di conciliare innovazione tecnologica e protezione ambientale occorre che l'IA venga messa al servizio della tutela dell'ambiente. L'obiettivo è quello di traslare il *focus* della ricerca scientifica da una *sustainability AI* a un *AI for sustainability*, antecedendo la tutela ambientale agli interessi economici delle *big tech*.

Bisogna inoltre superare le criticità legate al rinvio agli organismi di *standardizzazione*, implementando un maggiore coordinamento normativo tra le norme direttamente volte a tutelare l'ambiente e quelle volte a disciplinare l'IA.

L'auspicio è che le future normative si occupino di integrare i criteri della sostenibilità ambientale nella regolamentazione di queste tecnologie, introducendo per le aziende obblighi di trasparenza e di rendicontazione sulle emissioni e limiti al consumo energetico e idrico. In questi termini, si rende necessaria l'elaborazione di una legislazione *ad hoc* che intervenga per introdurre a disciplinare nuovi profili di responsabilità ambientale a carico delle aziende che sviluppino sistemi di IA, al fine, da un lato, di indirizzarne l'impiego lì dove ciò sia funzionale a incrementare la ricerca e a facilitare i processi produttivi di ciascuno Stato, dall'altro, di

⁷² Green Software Foundation, *The EU AI Act: Insight from the Green AI Committee 2024*, consultabile all'indirizzo: www.greensoftware.foundation/articles/the-eu-ai-act-insights-from-the-green-ai-committee.

arginarne l'utilizzo qualora lo stesso si presenti potenzialmente lesivo della tutela ambientale. Una normazione più integrata e un coordinamento efficace tra le politiche digitali e ambientali potrebbero rappresentare un passo decisivo verso una regolazione capace di coniugare innovazione e sostenibilità.



ANDREA CASTALDO*

INTELLIGENZA ARTIFICIALE E DIRITTI FONDAMENTALI. L'APPROCCIO "ANTROPOCENTRICO" TRA CONSIGLIO D'EUROPA E AI ACT

SOMMARIO: 1. Introduzione – 2. Convenzione del Consiglio d'Europa sull'AI e Regolamento (UE) 1689/2024: due facce della stessa medaglia? – 3. Brevi accenni al rapporto AI-Giudice – 4. Il possibile ruolo della Corte/Convenzione EDU – 5. Brevi considerazioni conclusive.

1. Introduzione

Come è noto l'Intelligenza Artificiale (AI)¹ ha assunto, negli ultimi anni, un rilievo assolutamente primario all'interno del dibattito giuridico, alla luce di una tensione tra le

* Dottorando di ricerca in Scienze giuridiche (curriculum internazionalistico-europeo-comparato) presso l'Università degli Studi di Salerno e Avvocato.

¹ Al singolare, così da intendersi come tecnologia informatica che mira a far sì che le macchine possano simulare l'intelligenza umana, ma ben consci che il progresso tecnologico ha portato all'affermarsi di più intelligenze artificiali, intesi come prodotti di mercato. Per una panoramica sul tema, *ex multis*, G. FINOCCHIARO, *Diritto dell'intelligenza artificiale*, Bologna, 2024; S. CORRADI, *Intelligenza artificiale e diritto: una prospettiva filosofica e giuridica*, Roma, 2024; C. GRIECO, *Intelligenza Artificiale e tutela degli utenti nel diritto dell'Unione europea*, Napoli, 2023; ID., *Intelligenza Artificiale e Diritti Umani nel Diritto Internazionale e dell'Unione europea*, in *Ordine internazionale e diritti umani* OIU, 2022, 3, p. 782ss; R. RAZZANTE, *Manuale sull'intelligenza artificiale*, Torino, 2024; E. FAZIO, *Intelligenza artificiale e diritti della persona*, Napoli, 2023; L. SUÁREZ FERNÁNDEZ, *La proposta di Regolamento dell'Unione europea sull'intelligenza artificiale tra diritti fondamentali, mercato e progresso tecnologico*, in *Tecnologie e diritto*, 2023, 2, p. 135ss.; A. PAJNO, F. DONATI, A. PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, I, Bologna, 2022; G. SARTOR, *L'intelligenza artificiale e il diritto*, Torino, 2022; M. TAMPIERI, *L'intelligenza artificiale e le sue evoluzioni*, Padova, 2022; G. FINOCCHIARO, *La Proposta di Regolamento sull'intelligenza artificiale*, in *Il diritto dell'informazione e dell'informatica*, 2022, 2, p. 303ss.; G. RESTA, *Cosa c'è di europeo nella Proposta di Regolamento UE sulla intelligenza artificiale*, *ivi*, p. 323ss.; A. F. URICCHIO, G. RICCIO, U. RUFFOLO (a cura di), *Intelligenza Artificiale tra etica e diritti*, Pisa, 2021; A. D'ALOIA (a cura di), *Intelligenza artificiale e diritto*, Milano, 2021; G. ALPA (a cura di), *Diritto e intelligenza artificiale*, Pisa, 2020; G. TADDEI ELMI, A. CONTALDO (a cura di), *Intelligenza artificiale*, Pisa, 2020; U. RUFFOLO (a cura di), *Intelligenza artificiale - il diritto, i diritti, l'etica*, Milano, 2020; S. DORIGO (a cura di), *Il ragionamento giuridico nell'era dell'intelligenza artificiale*, Pisa, 2020; P. PERLINGIERI, S. GIOVA, I. PRISCO (a cura di), *Il trattamento algoritmico dei dati tra etica, diritto ed economia*, Napoli, 2020; G. SARTOR, *Artificial intelligence and human rights: Between law and ethics*, in *Maastricht Journal of European and Comparative Law*, 2020, 6, p. 705ss.

molteplici possibilità di impattare, si spera con miglioramenti, sulle nostre vite e i mai sopiti timori riguardanti le nuove frontiere tecnologiche, considerando il crescente potere delle “macchine” innanzi alla fragilità e alla dignità della persona umana. A prescindere dall’ambizione, o dalla preoccupazione, di controllare pienamente questo fenomeno, il fatto che una macchina sia in grado di realizzare funzioni che fino ad ora erano proprie dell’uomo costituisce una sfida epocale, che segna la quarta rivoluzione industriale e un mutamento radicale della condizione umana².

Il suo potenziale e l’interazione con l’uomo influenza le dinamiche economiche, politiche e sociali e, conseguentemente, chiede alle scienze giuridiche una risposta che consenta di adeguare i rapporti sociali al nuovo contesto³.

L’AI è passata in pochi anni da essere tema futuristico e d’avanguardia ad argomento di «(ab)uso corrente»⁴. Si addivene all’affermazione dell’“impersonale”, e la nozione stessa di “persona” viene messa in discussione.

«Operare senza regole è il più faticoso e difficile mestiere di questo mondo» come diceva il Manzoni⁵ e lo è altresì senza una definizione del concetto di Intelligenza Artificiale.

Infatti, è utile delinearne il significato per evitare di alimentare una deriva comune dei nostri giorni e cioè quella di discutere senza aver un comune dizionario; questo in quanto bisogna chiarire come sia in realtà complesso fornire una definizione univoca di AI: non può dirsi che ne esista una generalmente riconosciuta, ciò per via dell’elevato rigore tecnico della materia, del progresso tecnologico e dell’interdisciplinarietà⁶.

In primis, è stato il Parlamento europeo ad aver avuto il merito di aver fornito una definizione il più possibile eterogenea e puntuale, intendendo l’AI come l’abilità di una “macchina” di mostrare capacità umane quali il ragionamento, l’apprendimento, la pianificazione e la creatività. L’AI si mette in relazione con quello che percepisce e risolve problemi, agendo verso un obiettivo specifico, altresì è capace di adattare il proprio comportamento analizzando gli effetti delle azioni precedenti e lavorando in autonomia⁷.

Si veda anche l’art. 3 dell’*AI Act*⁸ che definisce il sistema di AI 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, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali».

Il rapporto tra AI e rispetto dei diritti fondamentali è divenuto, dunque, un tema imprescindibile; in particolare, nell’ottica europea tale rapporto è argomento complesso e contraddistinto da una rapida evoluzione. A tal uopo, scopo di questo scritto, nei limiti del

² U. RUFFOLO, *Prefazione*, in U. RUFFOLO, G. RICCIO, A. F. URICCHIO (a cura di), *Intelligenza artificiale tra etica e diritti*, Bari, 2021, p. 9.

³ La crescente presenza degli assistenti AI nella vita quotidiana può influenzare il nostro agire, creare dipendenza o manipolare gli utenti, sollevando interrogativi sull’agire umano, sulla dignità umana, minando l’autonomia individuale.

⁴ U. RUFFOLO, *Per i fondamenti di un diritto della robotica self-learning*, in ID. (a cura di), *Intelligenza artificiale e responsabilità*, Milano, 2017.

⁵ A. MANZONI, *Storia della colonna infame*, Milano, 2002.

⁶ C. CASONATO, *Costituzione e intelligenza artificiale: un’agenda per il prossimo futuro*, in *BioLaw Journal* (online), 2019, p. 712; P. COMOGLIO, *Spunti minimi per una definizione (normativa) di Intelligenza Artificiale alla luce della proposta di Regolamento UE del 21 aprile 2021*, in *Corti fiorentine*, 2021, p. 19ss.

⁷ Così il PARLAMENTO EUROPEO, *Che cos’è l’intelligenza artificiale?*, 3/9/2020.

⁸ Regolamento (UE) 2024/1689, del 13 giugno 2024, *che stabilisce norme armonizzate sull’Intelligenza artificiale*, in GUUE L del 12/7/2024.

breve spazio a disposizione, è fornirne una panoramica, senza alcuna pretesa di esaustività, in particolare alla luce della Convenzione Quadro del Consiglio d'Europa e dell'*AI Act*⁹, strumenti adottati di recente ma i cui effetti si vedranno nel medio-lungo periodo.

2. *Convenzione del Consiglio d'Europa sull'AI e Regolamento (UE) 1689/2024: due facce della stessa medaglia?*

Considerando la parziale coincidenza soggettiva tra la “piccola” (a 27 Stati) e “grande” (a 46 Stati) Europa, i due presidi principali ai quali è affidata la tutela dei diritti fondamentali sono costituiti dalla Carta dei diritti fondamentali dell'Unione europea (Carta di Nizza/CDFUE) e dalla Convenzione per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali (CEDU).

Una vera protezione dei diritti fondamentali contro le eventuali distorsioni dell'AI può essere conseguita soltanto se vi è una collaborazione dell'Unione europea con le altre organizzazioni internazionali¹⁰, tra cui appunto il Consiglio d'Europa, «al fine di tutelare l'integrità del diritto dell'Unione e di garantire che le norme del diritto internazionale e unionale rimangano coerenti»¹¹. Il tutto nel quadro di una «integrazione delle tutele», così che il rispetto degli obblighi internazionali possa costituire uno «strumento efficace di ampliamento della tutela stessa», mirando alla «massima espansione delle garanzie»¹².

Lo scopo degli strumenti sovranazionali adottati da Consiglio d'Europa e Unione europea è quello di garantire la tutela dei diritti fondamentali e dei principi di libertà e di democrazia su cui si fonda lo Stato costituzionale¹³. Di pari passo con la ricerca di una definizione, la riflessione giuridica, sempre più presente in questo settore, ha mostrato una

⁹ In via non esaustiva, v. AA.VV., *Navigare l'European AI Act*, Milano 2024; G. TADDEI ELMI, A. CONTALDO, *AI Act - Regolamento (UE) 1689/2024*, Pisa, 2024; M. BORGOBELLO, *AI Act. Il regolamento europeo sull'Intelligenza artificiale*, Milano, 2024.

¹⁰ Si pensi all'*High-level Advisory Body on Artificial Intelligence* dell'ONU che ha presentato un rapporto per una *governance* globale dell'AI e alla Risoluzione dell'Assemblea Generale delle Nazioni Unite A/78/L49, 11/3/2024 in cui si sottolinea il rispetto, la tutela e la promozione dei diritti umani nella progettazione, nello sviluppo, nell'implementazione e nell'uso dell'AI. Cfr. G. ZACCARONI, *Intelligenza artificiale e principio democratico: riflessioni a margine dell'emersione di un quadro normativo europeo*, in F. FERRI (a cura di), *L'Unione europea e la nuova disciplina sull'intelligenza artificiale: questioni e prospettive*, Quaderni AISDUE (online), 2024, 2, pp. 19-50.

¹¹ Così al Considerando 7 della Decisione (UE) 2022/2349 del Consiglio, del 21 novembre 2022, *che autorizza l'avvio di negoziati a nome dell'Unione europea per una convenzione del Consiglio d'Europa sull'intelligenza artificiale, i diritti umani, la democrazia e lo Stato di diritto*, in GUUE L 311 del 2/12/2022.

¹² B. NASCIBENE, *La tutela dei diritti fondamentali in Europa: i cataloghi e gli strumenti a disposizione dei giudici nazionali*, in SCUOLA SUPERIORE DELLA MAGISTRATURA (a cura di), *I diritti fondamentali fra Carte e Costituzioni europee*, Roma, 2022, p. 19 s. Cfr. I. ANRÒ, *Carta dei diritti fondamentali dell'Unione europea e CEDU: dieci anni di convivenza*, in *Federalismi.it* (online), 2020, p. 109 s.; A. DI STASI, *CEDU e ordinamento italiano*, Padova, 2020; ID., *Lo spazio europeo di giustizia come segmento del più ampio spazio europeo di libertà, sicurezza e giustizia: incidenza virtuosa della tutela dei diritti fondamentali e limiti alla sua realizzazione*, in A. DI STASI, A. IERMANO, A. LANG, A. ORIOLO, R. PALLADINO, *Spazio europeo di giustizia e applicazione giurisprudenziale del Titolo VI della Carta dei diritti fondamentali dell'Unione europea*, Napoli, p. 13 s.; ID., *Introduzione alla Convenzione Europea dei Diritti dell'Uomo e delle Libertà Fondamentali*, Padova, 2025; ID., *La Convenzione europea dei diritti umani: l'effettività di un unicum a 70 anni dalla sua firma*, in *Freedom, Security & Justice FSJ* (online), 2020, 3, pp. 1-13.

¹³ V. J. ZILLER, *The Council of Europe Framework Convention on Artificial Intelligence vs. The EU Regulation: Two Quite Different Legal Instruments*, in *Rivista interdisciplinare sul Diritto delle Amministrazioni Pubbliche* (online), 2024, 2 p. 202-227.

certa difficoltà a concepire una regolazione di più ampio respiro, anche in virtù della mancanza di una definizione uniforme di AI.

Numerose sono le preoccupazioni scaturite dallo sviluppo e dall'implementazione dell'AI – un'infrastruttura cognitiva globale – che, se in alcuni casi è ai primi utilizzi, già dimostra grandi abilità, preoccupanti conseguenze e sviluppi potenzialmente inimmaginabili¹⁴.

In tal senso, la Commissione europea¹⁵, ha evidenziato come «dato l'impatto significativo che l'AI può avere sulla nostra società e la necessità di creare maggiore fiducia, è essenziale che l'AI europea sia fondata sui nostri valori e diritti fondamentali quali la dignità umana e la tutela della privacy»¹⁶.

Ancora, sempre la Commissione ha sottolineato che «l'uso dell'AI può pregiudicare i valori su cui si fonda l'Unione e causare violazioni dei diritti fondamentali, compresi i diritti alle libertà di espressione e di riunione, la dignità umana, la non discriminazione (...) la protezione dei dati personali e della vita privata o il diritto a un ricorso giurisdizionale effettivo e a un giudice imparziale, nonché la tutela dei consumatori (...). Le caratteristiche specifiche di molte tecnologie di AI, tra cui l'opacità, la complessità, l'imprevedibilità e un comportamento parzialmente autonomo, possono rendere difficile verificare il rispetto delle normative dell'UE in vigore volte a proteggere i diritti fondamentali e possono ostacolarne l'applicazione effettiva»¹⁷.

L'innovazione tecnologica è la leva del progresso e della crescita di società e Stati, ma deve sempre mantenere un saldo ancoraggio etico e coniugare la libertà con la responsabilità, senza lasciare indietro nessuno¹⁸ e promuovendo il benessere individuale e collettivo¹⁹.

Orbene la Convenzione del Consiglio d'Europa²⁰ costituisce un corposo testo giuridico con un approccio umano-centrico. La sua potenziale, ampia applicazione e la centralità del rispetto dei diritti umani potrebbero fornire un contributo per porre un freno a uno sviluppo tecnologico che presenta notevoli rischi per principi e valori democratici. Il Rapporto

¹⁴ Cfr. O. POLLICINO, P. DUNN, *Intelligenza artificiale e democrazia*, Milano, 2024.

¹⁵ COMMISSIONE EUROPEA, *Libro Bianco sull'Intelligenza artificiale. Un approccio europeo all'eccellenza e alla fiducia*, Com(2020)65, 19/2/2020.

¹⁶ Per una attenta riflessione sulla tutela dei diritti nello spazio digitale europeo, v. F. BUONOMENNA, *Brevi riflessioni sulla tutela dei diritti nello "spazio digitale" europeo*, in *Freedom, Security & Justice FSJ* (online), 2025, 2, pp. 228-241.

¹⁷ ID., *Libro Bianco*, cit., p. 12. Vds. Studio del CoE sull'uso dell'IA e delle possibili ripercussioni sui diritti fondamentali e Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, *L'Intelligenza artificiale per l'Europa*, Com(2018)237, 15/4/2018. Tali preoccupazioni della Commissione europea in merito ai diritti fondamentali emergono sin dal primo considerando dell'*AI Act*.

¹⁸ Lo scopo è evitare casi di cd. "esclusione digitale", v. F. CAVALLO, M. PIERANI, *Anziani e poveri esclusi dal digitale, il dramma ignorato dal Governo*, in *Agenda Digitale* (online), 13/12/2021.

¹⁹ Cfr. R. RAZZANTE, *AI e tutela dei diritti fondamentali*, su *Dirittifondamentali.it* (online), 2024, 1.

²⁰ COE, *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, CM(2024)52, 17/05/2024. Si è avuto il coinvolgimento di 46 Stati membri del Consiglio d'Europa, dell'Unione europea e di 11 Stati non membri. L'Italia non solo non l'ha ratificata (nessuno dei firmatari lo ha fatto, ad oggi), ma neanche firmata. Altresì, rileva il richiamo a precedenti Atti, quali la Raccomandazione sull'intelligenza artificiale adottata dal Consiglio dell'OCSE, la Raccomandazione del Comitato dei Ministri del Consiglio d'Europa agli Stati membri sugli impatti dei sistemi algoritmici sui diritti umani, la Raccomandazione UNESCO sull'etica dell'intelligenza artificiale e lo stesso *AI Act*. Si rimanda per un approfondimento delle disposizioni ad A. IERMANO, *Consiglio d'Europa e Intelligenza Artificiale: Un primo tentativo di regolamentazione a tutela di diritti umani, democrazia e Stato di Diritto*, in *Freedom, Security & Justice FSJ* (online), 2025, 2, p. 30 s.

esplicativo²¹ evidenzia come l'obiettivo è quello di avere «un quadro giuridico comune a livello globale al fine di applicare gli obblighi giuridici nazionali e internazionali esistenti (...) nella sfera dei diritti umani, della democrazia e dello Stato di diritto»²², in relazione a qualsivoglia attività realizzata dall'AI.

È particolarmente rilevante che l'approccio adottato fa sì che l'Atto si concentri sugli aspetti attinenti alla promozione e alla protezione dei diritti umani, tralasciando invece la regolazione degli aspetti economici e di mercato dell'AI, prospettiva invece adottata nell'*AI Act* dell'UE²³.

Come il Regolamento europeo, condivide però un *risk-based approach* con specifiche disposizioni per le valutazioni del rischio e d'impatto e misure per la mitigazione del rischio.

La definizione di AI alla base della Convenzione *de qua* è la medesima utilizzata nell'*AI Act*²⁴, infatti, entrambi gli Atti si rifanno alla definizione adottata dall'OCSE²⁵ con l'idea di potenziare la cooperazione della *governance* tecnologica a livello globale²⁶.

Si evidenzia però la genericità dei suoi obblighi e l'ampio margine di discrezionalità lasciato alle Alte Parti contraenti, il che implica che il suo impatto rischia di rivelarsi limitato, in particolar modo nel coinvolgere le multinazionali della tecnologia²⁷, vere destinatarie di tali discipline.

La Commissione europea ha firmato tale Convenzione a nome dell'Unione europea, stabilendo quindi una connessione con il Regolamento sull'AI²⁸, e sancendo la volontà di diventarne parte in attesa dell'intervento di Consiglio e Parlamento.

²¹ COMMITTEE ON ARTIFICIAL INTELLIGENCE, *Explanatory Report, Framework Convention on artificial intelligence, human rights, democracy and the rule of law*, CM(2024)52, 17/5/2024.

²² Si vedano ad esempio gli artt. 5, 7, 8, 10 e 11 della Convenzione.

²³ L'*AI Act* vero che contiene disposizioni volte a tutelare i diritti fondamentali, offrendo un livello di tutela complementare a quello previsto da altri strumenti legislativi, ma per tutti i restanti diritti, vi è solo un riferimento altrettanto generico alla Carta dei diritti fondamentali dell'Unione europea. Per approfondimenti v. P. DE PASQUALE, *The European Declaration on Digital Rights and Principles: Towards a Human-Centred Digital Constitutionalism*, in C. SCHEPISI, V. CAPUANO, S. LATTANZI (eds.), *An Introduction to Digital Data Law in the EU: Regulatory framework and future perspectives*, Torino, 2025.

²⁴ Si è scelta una definizione diversa rispetto a quella, ampia e tecnica, del Parlamento UE che si riferiva ad una AI idonea a "simulare"/"imitare" l'intelligenza umana, vds. Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale (2020/2014(INL)), Risoluzione del Parlamento europeo del 20 gennaio 2021 sull'intelligenza artificiale (2020/2013(INI)) e G. PROIETTI, *Le definizioni di sistemi di intelligenza artificiale nelle proposte legislative europee. Un'analisi critica*, in *Dialoghi di Diritto dell'Economia*, 15/11/2023.

²⁵ OCSE, *Recommendation of the Council on Artificial Intelligence*, I, OECD/LEGAL/0449, 22/5/2019.

²⁶ Così come previsto anche dalla Bletchley Declaration by Countries Attending the AI Safety Summit.

²⁷ Ad oggi, l'ecosistema dell'AI è ad appannaggio esclusivo di un numero ristretto di privati localizzati tra Cina e USA che ne controllano la produzione, l'addestramento e la distribuzione, e ne influenzano la regolazione. Gli obblighi generali sanciti dalla Convenzione sono rivolti alla tutela dei diritti dell'uomo e all'integrità del processo democratico ed altresì alla protezione della *rule of law*. Non si dimentichi che ciò riflette anche i tre diversi approcci all'uso e alla regolamentazione delle tecnologie digitali, valori e principi fondamentali cui si ispirano i rispettivi ordinamenti: gli USA hanno dato vita ad un modello *market-driven*, la Cina ad uno *state-driven*, l'UE ad un modello *rights-driven*. Possiamo dire che gli USA di Trump puntino ad una *deregulation* sull'AI, con effetti che potrebbero riverberarsi anche nell'UE, come già visto dal "Digital Omnibus" in merito alla *governance* complessiva del diritto digitale nell'UE. Anche se l'*AI Act*, con i suoi obblighi stringenti, dovrebbe rappresentare un approccio opposto rispetto alla linea americana. Per approfondimenti v. G. PITRUZZELLA, *Il costituzionalismo digitale tra Stati Uniti e Europa*, in *Unione europea e diritti* (online), 2025, 2, pp. 1-11.

²⁸ La firma della Commissione UE alla Convenzione rappresenta un altro passo in avanti nel lungo percorso di avvicinamento, coordinamento e interazione tra UE e Consiglio d'Europa nel più ampio capitolo dell'adesione dell'UE alla CEDU, affiancandosi al dettato dell'art. 6, parr. 2 e 3, TUE e artt. 51 e 52 CDFUE. Cfr. A. DI

I legislatori, nazionali ed europei, inseguono fattispecie mutevoli e sfuggenti connesse all'utilizzo dell'AI, ma indiscutibilmente, il cd. *AI Act*, è un primo importante punto fermo dell'UE in tale settore. Proprio l'art. 1 di tale Atto pone l'uomo al centro, infatti, il suo scopo è quello di «promuovere la diffusione di un'Intelligenza Artificiale antropocentrica e affidabile», considerando anche un altro assunto focale e cioè che l'AI deve essere lecita ed etica²⁹.

Di qui la centralità del diritto nell'analisi delle strumentazioni AI, in quanto alla base delle problematiche e criticità, che si legano all'uso di tali tecnologie, vi sono le lesioni e le distrazioni alle tutele fondamentali dell'essere umano.

Gli ambiti di utilizzo dell'AI in cui si può realizzare, in caso di uso distorto della tecnologia, una violazione e/o lesione dei diritti fondamentali sono molteplici. Sul punto, è intervenuta in anticipo, con acume e lungimiranza, l'*European Union Agency for Fundamental Rights* (FRA)³⁰, nel cui studio si individuano i rischi che si annidano nell'impiego dell'AI, concentrandosi su quattro ambiti sensibili, quali: servizi sociali, attività di polizia predittiva, diagnosi medica e pubblicità mirata.

Ma l'affermazione particolarmente rilevante riguarda la necessità di assicurare che l'AI rispetti tutti i diritti fondamentali – così come sanciti nella CEDU – e i valori alla base dell'UE – previsti nei Trattati, ad es. l'art. 2 TUE³¹, e nella CDFUE –, che vengono in rilievo a seconda del contesto e dell'uso che si fa dei sistemi di AI.

È possibile affermare, sulla scorta di quanto sottolineato anche in tale studio, come non vi sia la piena consapevolezza e, di conseguenza, la dovuta considerazione della totalità dei *fundamental/human rights* su cui l'AI è suscettibile di incidere³². I potenziali benefici che l'AI può avere, e sta avendo, in numerosi settori³³ e aspetti della nostra vita quotidiana sono di certo plurimi, ma bisogna continuare a perseguire un approccio antropocentrico, volto a garantire il rispetto della legislazione in materia di diritti fondamentali, mirando ad integrare

STASI, *Introduzione alla Convenzione*, cit., p. 44ss. e M. PARODI, *L'adesione dell'Unione Europea alla CEDU*, Napoli, 2020. L'eventuale futura ratifica della stessa da parte dell'UE farebbe sì che la Convenzione quadro diventi parte dell'*acquis* comunitario e l'applicazione dinanzi ai giudici degli Stati membri. Cfr. G. ZACCARONI, *Intelligenza artificiale e principio democratico*, cit., p. 21.

²⁹ Cfr. COMMISSIONE EUROPEA, *Orientamenti etici per un'AI affidabile*, 17/7/2020.

³⁰ FRA'S FOCUS PAPER, *Getting the Future Right, Artificial Intelligence and Fundamental Rights*, 2020, pp. 15 ss. e 47 ss.

³¹ Si fa riferimento a valori quali la dignità umana, la libertà, la democrazia, l'uguaglianza, lo Stato di diritto e i diritti umani, il pluralismo, la non discriminazione, la tolleranza, la giustizia, la solidarietà e la parità tra donne e uomini.

³² Per le enormi criticità, v. lo studio OPENAI, *Disrupting malicious uses of AI*, 5/5/2025.

³³ In verità, già oggi, è possibile coglierne gli aspetti positivi in vari ambiti: la prevenzione della disinformazione e degli attacchi informatici e l'accesso a informazioni di qualità possono contribuire a rafforzare la democrazia. L'AI poi supporta i medici nell'identificare le malattie, permettendo così di prolungare la vita umana. Si pensi a *MedGemma* o *MAI-DxO*, AI in grado di semplificare la documentazione clinica, suggerire diagnosi e cure, selezionare gli esami a cui sottoporre il paziente, svolgere ricerche in ambito medico, una potente estensione della capacità clinico-sanitarie del medico. Altresì, l'AI svolge un ruolo d'impatto nei sistemi educativi e formativi, anche con derive di non poco conto formulando test, quesiti o svolgendo temi al posto del discente. Non si può dimenticare l'impatto dell'AI nel settore giudiziario, nel pubblico impiego, nei servizi pubblici e nella sicurezza. In Italia, la Camera dei deputati ha presentato tre progetti basati sull'AI per assistere i deputati e gli uffici legislativi nell'attività di normazione. Ancora, da considerare il ruolo svolto nel settore dei trasporti e nella produzione di energia sostenibile, così come per la rilevazione di frodi nelle transazioni o truffe. In futuro potrebbe essere usata per prevedere e localizzare i terremoti o eventi climatici estremi. In campo militare, l'AI già è usata per la difesa e le strategie di attacco, anche alla luce dell'accordo tra Pentagono e OpenAI.

i requisiti di responsabilità e trasparenza nello sviluppo di AI cd. “ad alto rischio”³⁴. Ciò consente anche un miglioramento nel settore legislativo, in quanto si cerca di garantire che tali sistemi siano progettati, sin dall’inizio, tenendo conto del rispetto dei diritti fondamentali della persona. Qualora si verificassero delle violazioni, ad esempio, tali requisiti consentirebbero alle autorità nazionali di avere accesso alle informazioni necessarie per indagare se l’AI sia stata usata in modo conforme al diritto dell’Unione.

In questo senso la «Relazione sull’intelligenza artificiale in un’era digitale» del Parlamento europeo del 2022 sottolinea che l’*AI Act* «potrebbe dare all’UE una leva e il vantaggio della prima mossa nella definizione di norme internazionali sull’AI basate sui diritti fondamentali, nonché esportare con successo una “AI affidabile” antropocentrica in tutto il mondo», affermando una sovranità digitale europea nel mondo³⁵.

Per garantire il rispetto di tali presidi è anzitutto necessario che l’uso di tale tecnologia venga effettuato nel pieno rispetto della dignità umana, presupposto di tutti gli altri diritti fondamentali riconosciuti dall’ordinamento europeo quali, ad esempio, il principio di non discriminazione, il diritto alla sicurezza, i principi di trasparenza, imparzialità ed equità³⁶. Così come i diritti collettivi, potrebbero essere potenzialmente minacciati laddove l’AI, tramite sistemi di riconoscimento biometrico e facciale, venga utilizzata per ledere o minacciare la libertà di stampa, quella di riunione o di associazione, la democrazia, il diritto di voto, in generale, lo Stato di diritto posto alla base delle istituzioni democratiche nazionali e sovranazionali³⁷.

³⁴ La classificazione del rischio, così come sancita dall’*AI Act*, si basa sulla finalità, modalità, funzione svolta e sull’utilizzo finale. Alla luce di ciò, i sistemi AI possono essere classificati come ad alto rischio se il sistema di AI: 1) è integrato come componente di sicurezza in prodotti disciplinati dalla legislazione europea vigente in materia; 2) è destinato a essere utilizzato per un caso d’uso ad alto rischio, in settori quali l’istruzione, il lavoro, le attività di contrasto alla criminalità o nell’ambito dei fenomeni migratori.

³⁵ Cfr. A. BRADFORD, *Effetto Bruxelles*, Milano, 2021.

³⁶ R. RAZZANTE, *AI e tutela dei diritti fondamentali*, cit. e J. DE BRUYNE, C. VANLEENHOVE (eds.), *Artificial Intelligence and the Law*, Cambridge, 2022.

³⁷ C. SCHEPISI, *Le “dimensioni” della regolazione dell’Intelligenza Artificiale nella proposta di Regolamento della Commissione*, in *Quaderni AISDUE* (online), 2022, 16, p. 347ss. L’AI costituisce un serio rischio per la Democrazia laddove venga utilizzata per creare immagini, video e audio falsi ma estremamente realistici, cd. *deepfake*, usati per truffe, per danneggiare la reputazione altrui o alterare i processi decisionali, come accaduto di recente alle elezioni in Romania. Effetti negativi possono aversi sull’ambiente, sulla protezione dei dati e sul diritto alla vita privata oppure ulteriori danni stanno emergendo in quanto l’AI (e i suoi costruttori) viola la normativa sul diritto d’autore ai fini del suo addestramento. Il Rapporto dell’Alto Commissario delle Nazioni Unite per i Diritti Umani, *Il diritto alla privacy nell’era digitale*, Ginevra, 2021, nell’evidenziare l’impatto dell’AI sui diritti umani, come la *privacy*, la sicurezza nazionale e la giustizia, lanciava un drammatico appello. Ancora, l’AI, se da un lato è in grado di creare posti di lavoro, dall’altro, al momento molti ne sta togliendo, tra istituti finanziari e imprese, es: i circa dieci mila posti di lavoro tagliati da Microsoft o l’utilizzo sempre più massiccio da parte di Amazon di robot nei propri magazzini. Secondo l’*Artificial Intelligence Index Report 2025* l’AI ha superato le prestazioni umane in compiti tecnici in cui prima non era all’altezza, l’unica attività (per ora) in cui l’AI non ha ancora raggiunto gli umani è la comprensione e il ragionamento multimodale, poiché fondate su elaborazione e ragionamento. Senza dimenticare le preoccupazioni nel settore giustizia per l’utilizzo di sistemi AI di analisi delle prove o di giustizia predittiva con lo scopo di prevedere decisioni giudiziarie future sfruttando i precedenti giurisprudenziali, il tutto sull’idea che i precedenti giurisprudenziali, se analizzati su larga scala, abbiano una intrinseca regolarità e ciclicità che possono essere appresi da una macchina al fine di emettere essa stessa decisioni in luogo del giudice-umano. Sul punto, v. Relazione della Commissione al Parlamento europeo, al Consiglio, al Comitato Economico e Sociale europeo e al Comitato delle Regioni, *Tutela dei diritti fondamentali nell’era digitale - Relazione annuale 2021 sull’applicazione della Carta dei diritti fondamentali dell’Unione europea*, Bruxelles, COM(2021)819 def, 10/12/2021; G. CANZIO, *AI act e processo penale: sfide e opportunità*, in *Sistema Penale* (online), 14/10/2024; A. GIOVENE, *AI Act. Tra perfettibilità e compromessi*, in *Federalismi.it* (online), 17/4/2024; N. LIPARI,

3. Brevi accenni al rapporto AI-Giudice

Oggetto di normazione negli Atti summenzionati nel paragrafo che precede, nonché della legislazione nazionale *in fieri*, è il settore giustizia, nel quale già vi sono noti e vari utilizzi dell'AI.

Si vuole porre l'attenzione sul ruolo dei giudici, in quanto, considerando il Disegno di Legge attualmente in discussione in Parlamento³⁸, il magistrato potrà avvalersi dell'uso dell'AI anche per la stesura del provvedimento che definisce il giudizio ed è esonerato dall'obbligo di informare le parti. Infatti, l'art. 15 del DDL prevede che «Nei casi di impiego dei sistemi di intelligenza artificiale nell'attività giudiziaria è sempre riservata al magistrato ogni decisione sull'interpretazione e sull'applicazione della legge, sulla valutazione dei fatti e delle prove e sull'adozione dei provvedimenti»³⁹.

Rilevano innanzitutto gli artt. 47 e 48 CDFUE e il corrispondente art. 6 CEDU, altresì, sul punto nell'art. 6, par. 3, dell'*AI Act* letto in combinato disposto con l'Allegato III si trovano esplicitamente indicati «i sistemi di AI destinati a essere usati da un'autorità giudiziaria (...) nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti» come ad alto rischio «se presenta[no] un rischio significativo di danno per la salute, la sicurezza o i diritti fondamentali delle persone fisiche, anche nel senso di non influenzare materialmente il risultato del processo decisionale».

Alla luce poi del Considerando n. 61 dell'*AI Act*, si noti che «alcuni sistemi di AI destinati all'amministrazione della giustizia e ai processi democratici dovrebbero essere classificati come sistemi ad alto rischio, in considerazione del loro impatto potenzialmente significativo sulla democrazia, sullo Stato di diritto, sulle libertà individuali e sul diritto a un ricorso effettivo e a un giudice imparziale. È in particolare opportuno (...) classificare come ad alto rischio i sistemi di AI destinati a essere utilizzati da un'autorità giudiziaria (...) nelle attività di ricerca e interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti (...). L'utilizzo di strumenti di AI può fornire sostegno al potere decisionale dei giudici o all'indipendenza del potere giudiziario, ma non dovrebbe sostituirlo: il processo decisionale finale deve rimanere un'attività a guida umana».

Il legislatore europeo è preoccupato per lo svolgimento della funzione giudicante e sottolinea come essa rimanga prerogativa degli umani. D'altronde, l'*AI Act* impone una svolta nella *governance* algoritmica, così che l'innovazione tecnologica e la tutela giuridica diventano

Il diritto del nuovo millennio tra giurisdizionalizzazione ed algoritmo, in *Accademia*, 2024, p. 9 s.; A. SANTOSUOSSO, G. SARTOR, *Decidere con l'IA. Intelligenze artificiali e naturali nel diritto*, Bologna, 2024; A. FORMISANO, *L'impatto dell'intelligenza artificiale in ambito fondamentale giudiziario sui diritti*, su *Federalismi.it* (online), 11/9/2024 e R. GIUFFRÈ, P. PIAZZA, *Impiego dei sistemi di AI nell'attività giudiziaria: due pesi e due misure?*, 13/5/2025, in cui si evidenzia che «la giustizia predittiva si limita ancora, in buona sostanza, alla semplice continua selezione e riproposizione statistica di precedenti che più o meno si attagliano ad una situazione concreta, il che non assicura certo né la creatività giuridica né l'evoluzione del diritto, in quanto le rende dipendenti esclusivamente dalla quantità e qualità dei dati di apprendimento forniti al sistema, con conseguente appiattimento degli orientamenti giuridici».

³⁸ Si fa riferimento al DDL n. 1146, *Disposizioni e deleghe al Governo in materia di intelligenza artificiale*, approvato il 25/6/2025 con modificazioni dalla Camera dei deputati (Atto Camera n. 2316) e trasmesso al Senato della Repubblica. DDL e *AI Act* possono esser viste come due discipline complementari che agiscono per tutelare i diritti fondamentali nell'era dell'AI.

³⁹ Si ritiene scarsamente compatibile tale disposizione con la volontà antropocentrica degli atti europei nonché con la necessità che l'intero processo decisionale resti in capo alla persona fisica del giudice.

due facce della stessa medaglia, con l'obiettivo di addivenire alla creazione di un ecosistema digitale affidabile.

L'art. 5 della Convenzione del CoE evidenzia i rischi dell'AI nel «respect for the rule of law and the principle of access to justice and its proper administration; and respect for the principle of judicial independence». Ma altresì il fatto che l'AI non possa influenzare materialmente il risultato del processo decisionale è fonte di perplessità⁴⁰, comportando seri rischi per i diritti fondamentali. Il riferimento d'obbligo è al caso *Compas*⁴¹, riguardante il rischio di recidiva e l'influenza sul risultato decisionale avutosi, acuendo la discriminazione nei confronti di alcune categorie di imputati.

Parimenti, il problema dell'utilizzo dell'AI in tale delicato settore potenzialmente foriero di illimitati danni ai diritti fondamentali, non è tanto nei limiti al suo utilizzo, quanto nel rispetto del principio di trasparenza e quindi al tema della sua riconoscibilità, nell'*an* e nel *quomodo*, senza dimenticare i rischi di allucinazioni⁴².

4. Il possibile ruolo della Corte/Convenzione EDU

Mi sia permesso un riferimento finale alla Convenzione EDU⁴³, in quanto ritengo, visti i vari Stati aderenti, che questa debba (o possa) assumere il ruolo di «stella polare» verso le future sfide dell'AI, anche grazie ad una interpretazione evolutiva, sia dei principi nella stessa contenuti che della giurisprudenza della Corte, nonché a seguito di una concreta applicazione della Convenzione Quadro, così da poter avere sempre maggiori garanzie.

Allo stato attuale, non esiste ancora una giurisprudenza consolidata della Corte europea dei Diritti dell'Uomo (Corte EDU) avente ad oggetto le violazioni dei diritti umani causate dall'AI, come ad esempio, la responsabilità per una decisione discriminatoria presa da un algoritmo o la violazione della *privacy* da parte di un sistema di apprendimento automatico.

Sul punto vi sono *leading cases* aventi ad oggetto diritti fondamentali entrati in osmosi con l'AI, tra cui la protezione dei dati, la sorveglianza e i processi decisionali automatizzati, che è possibile considerare – in ottica preparatoria e preventiva – precursori di come la Corte EDU potrebbe pronunciarsi sull'AI.

Al riguardo degno di nota è il caso *S. e Marper v. UK*⁴⁴ sull'art. 8, con cui la Corte EDU ha condannato la conservazione indiscriminata di campioni genetici e ha stabilito principi

⁴⁰ Infatti, «la neutralità degli algoritmi è un mito, in quanto i loro creatori, consciamente o meno, riversano in essi i loro sistemi di valori», così Commissione Europea per l'efficienza della Giustizia, *Carta etica per l'uso dell'intelligenza artificiale nei sistemi giudiziari e nei loro ambiti connessi*, Strasburgo, 3/12/2018, p. 39. Cfr. A. SIMONCINI, *L'algoritmo incostituzionale: l'intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal* (online), 2019, 1, p. 69; G. RESTA, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, in *Politica del diritto*, 2019, 2, p. 214 e R. TREZZA, *La tutela della persona umana nell'era dell'intelligenza artificiale*, in *Federalismi.it* (online), 2022, 16, p. 277 s.; v. la successiva nota 46.

⁴¹ SUPREME COURT OF WISCONSIN, *State of Wisconsin v. Loomis*, n. 2015AP157-CR, 13/7/2016.

⁴² L. DENTIS, *Allucinazione da A.I. nell'atto difensivo e il nuovo art. 2407 c.c.*, su *Quotidiano Giuridico* (online), 4/4/2025; M. PAPPALARDO, *L'Intelligenza Artificiale sbaglia, l'avvocato paga*, *ivi*, 5/6/2025; L. CAPUTO, L. VIOLA, *L'IA non risponde della diffamazione causata dalle sue "allucinazioni"*, su *Domani* (online), 19/6/2025.

⁴³ Sul punto v. A. DI STASI, *Introduzione alla Convenzione Europea*, cit.

⁴⁴ Corte europea dei diritti dell'uomo, sentenza del 4 dicembre 2008, ricorso nn. 30562/04 e 30566/04, *S. and Marper v. the UK* [GC], ECLI:CE:ECHR:2008:1204JUD003056204.

chiave sul diritto al rispetto della vita privata che sono direttamente applicabili alla raccolta massiva di dati da parte dell'AI.

Ancora, sempre sullo stesso art. 8, il caso *Zakharov v. Russia*⁴⁵ che ha fissato criteri per la legalità dei programmi di sorveglianza delle comunicazioni, principi utilizzabili anche per valutare la legittimità dei sistemi di sorveglianza basati sull'AI.

Più di recente, il caso *Glukhin v. Russia*⁴⁶, sul tema del riconoscimento facciale, e del relativo potenziale impatto sulla libertà di espressione e di riunione (artt. 10 e 11).

Pro futuro, eventuali questioni potrebbero poi riguardare il *bias* algoritmico⁴⁷ e la discriminazione: il riferimento è all'art. 14 della Convenzione, e quindi a come poter garantire che le decisioni automatizzate dell'AI, in settori come l'assunzione, il *credit scoring*⁴⁸ o la valutazione del rischio di recidiva, non perpetuino e/o amplifichino pregiudizi esistenti⁴⁹.

Ancora, l'art. 6 prevede il diritto ad un processo equo e soprattutto alla motivazione della decisione⁵⁰, richiamando l'esempio del paragrafo precedente; in tal caso rileva come può assicurarsi che un soggetto comprenda (e possa contestare nel caso) efficacemente una decisione presa da un giudice⁵¹ con l'aiuto di un algoritmo o dal solo algoritmo, in quanto i sistemi di AI, già per gli stessi programmatori, sono oscuri nei processi decisionali e si parla appunto di *black box AI*⁵².

Infine, ci sarà da valutare la responsabilità e, quindi, come e a chi attribuire la responsabilità in caso di violazione dei diritti umani commessa da un'AI.

5. Conclusioni

In conclusione, i principi giuridici regolatori dell'AI possono essere tratti dalla Convenzione Quadro e dall'*AI Act* mentre la giurisprudenza esistente offre molti spunti per riflettere su come la Corte affronterà le cause future, riaffermando la centralità della Convenzione EDU nell'era dell'AI.

⁴⁵ Corte europea dei diritti dell'uomo, sentenza del 4 dicembre 2015, ricorso n. 47143/06, *Zakharov v. Russia* [GC], ECLI:CE:ECHR:2015:1204JUD004714306.

⁴⁶ Corte europea dei diritti dell'uomo, sentenza del 4 luglio 2023, ricorso n. 1519/20, *Glukhin v. Russia*, ECLI:CE:ECHR:2023:0704JUD001151920.

⁴⁷ Si fa riferimento a possibili distorsioni del processo decisionale o valutativo determinate da pregiudizi insiti all'interno di tali strumenti tecnologici.

⁴⁸ Corte di giustizia dell'Unione europea, sentenza del 27 febbraio 2025, causa C-203/22, ECLI:EU:C:2025:117.

⁴⁹ In particolare, sul rapporto AI e problematiche di genere v. A. DI STASI (a cura di), *Dalla non discriminazione alle Pari Opportunità. Intelligenza Artificiale e nuove Tecnologie, Sostenibilità e Gender Gap: il ruolo dell'Università. Atti del Convegno di studi tenuto in occasione della Giornata internazionale della donna 2024 - Università degli Studi di Salerno*, Milano, 2025.

⁵⁰ Il riferimento è ovviamente al diritto ad un equo processo, il quale include l'obbligo di motivazione delle sentenze. Quindi le decisioni giudiziarie, sia civili che penali, devono essere adeguatamente motivate, spiegando le ragioni che hanno portato alla decisione, sia in fatto che in diritto. La motivazione deve essere chiara, logica e sufficiente a consentire alle parti di comprendere le ragioni della decisione e, se del caso, di impugnarla.

⁵¹ Il problema sorge in particolar modo nel settore giudiziario e quindi coinvolge *prima facie* i giudici in quanto il magistrato potrà avvalersi dell'uso di strumenti di IA anche per la stesura del provvedimento che definisce il giudizio e, dall'altro, che è esonerato dall'obbligo di informare le parti così come richiesto per le professioni intellettuali.

⁵² Su tutti v. F. PASQUALE, *The Black Box Society*, Cambridge, 2015.

Tale novità legislativa deve confrontarsi con l'incognita della sua adeguatezza a un quadro tecnologico in rapidissima evoluzione. I confini definiti con gli strumenti regolatori sovranazionali saranno in grado di assicurare uno sviluppo “umanocentrico” dell'AI o il progresso tecnologico disorienterà le Istituzioni ponendo nuove criticità?

L'antropocentrismo, indicato come valore guida, non può e non deve, anche per motivi di sicurezza e protezione dei valori fondamentali di ogni Stato democratico – dalla non discriminazione, alla dignità dell'uomo, alla *privacy* e così via –, restare solo sulla carta.

L'approccio “*rights-driven*” scelto fa sì che la fiducia nell'AI non dipende solo dai risultati tecnologici, ma anche dalla capacità di garantire il rispetto dei fondamentali principi etici e giuridici⁵³.

Lo sviluppo dell'AI deve costituire un progetto europeo e sovranazionale, non nazionale. È un'opportunità che l'Europa, collettivamente, non può esitare a cogliere con fermezza. C'è bisogno di un dibattito aperto e inclusivo che coinvolga tutti gli attori, pubblici e privati, focalizzandosi sul modo più giusto di utilizzare queste nuove tecnologie e su come rispettare i diritti fondamentali⁵⁴.

Il ruolo del legislatore europeo, nel governo del processo di transizione digitale della società, è centrale, da una parte, per agevolare lo sviluppo di quello che sarà certamente un volano dell'economia moderna nei prossimi anni e, dell'altra, per fornire un'adeguata tutela dei diritti degli utenti che si trovino ad interagire con questi sistemi⁵⁵.

Giova richiamare, in ultimo, l'idea di “affermare” nuovi diritti al fine di mantenere l'umanocentrismo dell'AI: il diritto a non essere oggetto di misurazioni, analisi e addestramento e il diritto a un “contatto umano significativo”, tale da poter stabilire e sviluppare relazioni profonde con altri esseri umani⁵⁶.

Il Diritto, nel caso di fenomeni tanto dirompenti, non può solo seguire le evoluzioni tecnologiche ma deve pre-vederle⁵⁷.

⁵³ Cfr. A. BRADFORD, *Digital Empires*, Oxford, 2023 e F. DONATI, *La protezione dei diritti fondamentali nel Regolamento sull'Intelligenza Artificiale*, in O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI (a cura di), *La disciplina dell'Intelligenza Artificiale*, Milano, 2025.

⁵⁴ Così Ansip, già Vicepresidente della Commissione UE, alla presentazione del Libro Bianco - L'Intelligenza Artificiale a servizio del cittadino, a cura della *task force IA* dell'Agenzia per l'Italia digitale.

⁵⁵ Sul punto, decine di multinazionali hanno chiesto alla Commissione UE di rinviare l'entrata in vigore dell'*AI Act*, in virtù di norme poco chiare, complesse e sovrapponibili.

⁵⁶ R. VAN EST, J. GERRITSEN, L. KOOL, *Report Human rights in the robot age, Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality*, The Hague, 2017 e G. DE GREGORIO, *The rise of digital constitutionalism in the European Union*, in *International Journal of Constitutional Law*, 2021, 1, pp. 41-70.

⁵⁷ Vds. U. RUFFOLO, A. AMIDEI, *Intelligenza artificiale e diritti umani*, in *Diritti dell'Uomo*, 2020, 2, pp. 361-384 e R. STRAND, M. KAISER, *Report on Ethical Issues raised by emerging sciences and technologies*, Bergen, 2015.



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 innovate 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 sopraffatto 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 rinvencono 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. BATHAEE, *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) sottoponendo 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 accertare, prevenire, indagare e perseguire reati, a meno che tali sistemi non siano a disposizione del pubblico per segnalare un reato». Ed ancora, a riconferma 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 accertare, prevenire, indagare e perseguire reati o se è necessario per l'esercizio del diritto alla libertà di espressione e del diritto alla libertà delle arti e delle scienze garantito dalla Carta dei diritti fondamentali dell'UE, e fatte salve le tutele adeguate per i diritti e le libertà dei terzi».

3. 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 dallo *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'assurer 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 «les 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, notevole *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 precui 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 «significant» – 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 precetto 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 rinvencono, 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, *Vienpoint: 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 simile 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 eguaglianza 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 nonchè 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. BORGESIU, *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. BORGESIU, *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'erompere dell'intelligenza artificiale per lo sviluppo della pubblica amministrazione e dei servizi al cittadino*, in G.C. FERONI, C. FONTANA, E.C. RAFFIOTTA, *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 E 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 nasconda 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. SANTOSUOSSO, *Intelligenza artificiale, conoscenze neuroscientifiche 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 – comparando, 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).



ISABELLA SALSANO*

ARTIFICIAL INTELLIGENCE AND THE PROTECTION OF CULTURAL HERITAGE: HUMAN RIGHTS AND THE ROLE OF INTERNATIONAL LAW IN ADDRESSING FUTURE ETHICAL DILEMMAS

TABLE OF CONTENTS: 1. Introduction. – 2. The International Legal Framework for the Protection of Cultural Heritage. – 2.1. Cultural Heritage and Human Rights. – 3. Cultural heritage and human rights in the digital age. – 4. Ethical dilemmas and normative frictions in AI-driven cultural heritage practices. – 5. From cultural erosion to legal recognition: towards a right to collective memory in digital heritage governance. – 6. Conclusions.

1. *Introduction*

Artificial intelligence has been a paradigm-changing force for many domains, however, one of its most under-explored implications concerns its intersection with human rights and cultural heritage. AI technologies present enormous potential and urgent normative challenges, as they increasingly become part of the systems that manage cultural preservation, interpretation and access. The interaction between AI-mediated heritage practices and the international cultural-rights framework, rooted in Art. 27 of the Universal Declaration of Human Rights¹ and Art. 15 of the International Covenant on Economic, Social and Cultural Rights², underscores the growing recognition of heritage as a dimension of human dignity.

Such normative instruments support the right to engage in cultural life and the corresponding duty of governments to ensure equitable access and preservation, that - alongside participation - are mutually reinforcing components of the indivisible matrix of

*PhD student in International Law, University of Salento.

¹ Recognises the right to freely participate in cultural life, enjoy the arts, and share in scientific advancement. This implies the preservation and transmission of cultural knowledge and practices, which are essential components of collective memory. From now on it will be used the acronym UDHR.

² Affirms the right of everyone to take part in cultural life and obliges states to promote the conservation and development of culture. From now on it will be used the acronym ICESCR.

cultural human rights³. In this context, core research questions arise: to what extent can international legal frameworks ensure ethical governance and human rights protection in AI-driven practices concerning cultural heritage, and how can they safeguard collective memory against the risks of digital dispossession? This is especially relevant given the role AI plays in determining how society is accessed and interpreted: in the field of tangible heritage it allows restoration initiatives, real-time object recognition and virtual reconstructions of sites either destroyed or inaccessible⁴ and for intangible heritage, AI enables voice preservation, dialect analysis and the transmission of oral traditions through NLP⁵ and machine learning⁶.

These are presented as inherently democratising tools: 3D modelling, augmented reality and intelligent retrieval systems promise to broaden access to cultural artefacts for diverse populations, including those with disabilities or in remote areas⁷.

Yet digitisation relies on classificatory choices that may not preserve cultural context.

The translation of cultural expressions into data categories may produce representations that diverge from community understandings, thereby affecting the integrity and protection of heritage under international cultural rights frameworks⁸.

Moreover, AI-generated heritage may seem neutral while it often incorporates institutional perspectives, developer preferences or imaginative elements that distort the original meaning. In this sense, international legal systems are insufficient.

The 2021 UNESCO Recommendation on the Ethics of Artificial Intelligence affirms human rights, cultural diversity and inclusiveness as core principles guiding AI development.

However, as a non-binding instrument, it provides ethical guidance rather than enforceable obligations and contains no specific provisions addressing the protection of cultural heritage⁹. The AI Act of the European Union, on the other hand, focuses primarily

³ Y. DONDEES, Protection and promotion of cultural heritage and human rights through international treaties: two worlds of difference?, in C. WAELDE, C. CUMMINGS, M. PAVIS, H. ENRIGHT (eds.), *Research Handbook on Contemporary Intangible Cultural Heritage*, Cheltenham, 2018, pp. 54-77.

⁴ See, among the others, J. LI, *Application of artificial intelligence in cultural heritage protection*, in *Journal of Physics: Conference Series*, Vol. 1881, N. 3, 2021, pp. 1-7; B. R. DAS, H. B. MARINGANTI, N. S. DASH, Role of artificial intelligence in preservation of culture and heritage, in D. MISHRA, S. M. SAMANTA (eds.), *Digitalization of Culture through Technology*, London-New York, 2022, pp. 92-97; D. HARISANTY, K. L. B. OBILLE, N. E. V. ANNA, E. PURWANTI, F. RETRIALISCA, *Cultural heritage preservation in the digital age, harnessing artificial intelligence for the future: a bibliometric analysis*, in *Digital Library Perspectives*, 2024, vol. 40, pp. 609-630.

⁵ Referring to natural language processing.

⁶ Y. FU, K. SHI, L. XI, *Artificial intelligence and machine learning in the preservation and innovation of intangible cultural heritage: ethical considerations and design frameworks*, in *Digital Scholarship in the Humanities*, 2025, vol. 40, pp. 487-508; A. TSATSANASHVILI, *Artificial Intelligence in the Protection of Intangible Cultural Heritage*, in *European Journal of Transformation Studies*, 2024, pp. 163-178.

⁷ H. I. M. GAUDÊNCIO, *Culture Access and Technology: How Can Technology Democratize Cultural Access?*, Master thesis, Universidade NOVA de Lisboa, 2019.

⁸ M. GASPAROTTO, *Digital Colonization and Virtual Indigeneity: Indigenous Knowledge and Algorithm Bias*, Rutgers University, 2016, pp. 1-21.

⁹ B. S. SOLOMON, *UNESCO Best Practice Submission: A Unified Moral AI System for Global Ethics*, IT Analysis and Training Ltd., 2025; B. GIOVANOLA, P. GRANATA, Ethics for human-centered education in the age of AI, in F. SPIGARELLI, L. KEMPTON, L. CAMPAGNUCCI (eds.), *Entrepreneurship and Digital Humanities*, Cheltenham-Northampton, 2024, pp. 96-109; E. BEAN, C. BURLEIGH, C. HASKELL, T. BURRIS-MELVILLE, J. PAYNE, B. PATHAK, *Eavesdropping on UNESCO AI Policy, Leadership, and Ethics*, in *Journal of Leadership Studies*, 2025, vol. 18, n. 4, pp. 98-110.

on risk-based regulation, but has not so far established specific protections for intangible cultural heritage, that derives its meaning from community-based relational practices¹⁰.

AI systems that ignore these specificities can decontextualise traditions or introduce externally imposed standards of authenticity, resulting in forms of digital dispossession where cultural data are extracted or repurposed without the free, prior and informed consent of the communities concerned¹¹. This article addresses these tensions through a human rights lens, arguing that the preservation of heritage in the age of AI should be redefined as a moral and legal activity necessary to preserving the rights of people to access and share their legacy on their own conditions. This analysis is grounded in the international legal framework governing cultural heritage and draws on the human-rights-based approach developed in contemporary doctrine, outlining the normative bases of cultural heritage protection within international human rights law before examining the opportunities and risks associated with the use of AI in heritage contexts¹² and attempts to contribute to a fairer approach to the control of AI for cultural legacy.

2. *The international legal framework for the protection of cultural heritage*

The contemporary international framework for the protection of cultural heritage is the outcome of a gradual normative consolidation¹³. Initially conceived as an object of State sovereignty and national patrimony, cultural heritage progressively acquired an international legal dimension as the international community recognised its transboundary and universal value¹⁴. This shift was driven by the impact of armed conflicts and increasing awareness that heritage loss, whether through destruction, illicit trafficking or neglect, affects not only individual States but humanity as a whole¹⁵. From this evolution emerged a distinct *corpus* of rules, later complemented by human-rights developments, which frames cultural heritage as a protected interest requiring both domestic implementation and international cooperation¹⁶.

¹⁰ P. HACKER, *AI regulation in Europe: from the AI act to future regulatory challenges*, in *arXiv preprint*, 2023, pp. 1-18; I. KUSCHE, *Possible harms of artificial intelligence and the EU AI act: fundamental rights and risk*, in *Journal of Risk Research*, 2024, pp. 1-14.

¹¹ M. PISONI, G. GALENA ET AL., *Human-centred artificial intelligence for designing accessible cultural heritage*, in *Applied Sciences*, 2021, vol. 11, n. 2, p. 870.

¹² N. AHMAD, A. W. ALI, M. H. B. YUSSOF, *The challenges of human rights in the era of artificial intelligence*, in *UUM Journal of Legal Studies*, 2025, vol. 16, pp. 150-169.

¹³ L. V. PROTT, *Unesco International Framework for the Protection of the Cultural Heritage*, in J. A. R. NAFZIGER, A. M. NICGORSKI (eds.), *Cultural Heritage Issues: The Legacy of Conquest, Colonization and Commerce*, Leiden, 2009, pp. 257-285.

¹⁴ C. FORREST, *International Law and the Protection of Cultural Heritage*, London, 2010.

¹⁵ J. BLAKE, *On defining the cultural heritage*, in *International & Comparative Law Quarterly*, vol. 49, 2000, pp. 61-85. See also E. J. TECHERA, *Protection of Cultural Heritage Times of Armed Conflict: The International Legal Framework Revisited*, in *Macquarie Journal of International and Comparative Environmental Law*, 2007, pp. 1-20; J. HLADÍK, *Different legal issues related to the protection of cultural property in peacetime and wartime*, in *Proceedings of the ASIL Annual Meeting*, Vol. 106, 2012, pp. 453-462; P. GERSTENBLITH, *Archaeology in the context of war: Legal frameworks for protecting cultural heritage during armed conflict*, in *Archaeologies*, vol. 5, 2009, pp. 18-31.

¹⁶ M. PHELAN, *A synopsis of the laws protecting our cultural heritage*, in *New England Law Review*, vol. 28, 1993, pp. 63-108. J. H. MERRYMAN, "Protection" of the Cultural "Heritage"?, in *The American Journal of Comparative Law*, vol. 38, 1990, pp. 513-522.

The international protection of cultural heritage is therefore based on a normative architecture that integrates treaty law, soft-law instruments and human-rights jurisprudence.

The 1954 Hague Convention and its 1999 Second Protocol establish the first comprehensive regime, with obligations of safeguarding during armed conflict, including measures of prevention, such as the Enhanced Protection regime and individual criminal responsibility¹⁷. The Second Protocol also strengthens enforcement by establishing sanctions for grave breaches and increasing obligations of domestic implementation¹⁸.

It further states that the protection of cultural heritage is an obligation *erga omnes*, which reinforces the principle that it does not fall within the domain of domestic jurisdiction alone, even in cases of armed conflict¹⁹. Complementing conflict settings regulation, the 1970 UNESCO Convention deals with illicit import, export, and transfer of cultural property by establishing due-diligence duties, restitution mechanisms, and cooperation obligations²⁰.

It also operationalises the principle of good faith in restitution processes, requiring signatories to adopt appropriate legislative measures and museum-sector standards to prevent illicit circulation. A parallel shift occurred with the 1972 World Heritage Convention, which introduced the concept of Outstanding Universal Value and linked heritage protection to State obligations of identification and periodic reporting, establishing a system of international monitoring and collective oversight over State compliance²¹.

The 2003 Convention for the Safeguarding of the Intangible Cultural Heritage²² marked a decisive turn toward community-based protection, recognising that the transmission of intangible heritage depends on the consent of the communities concerned, and imposing on States the duty to ensure that safeguarding measures are community-driven and culturally appropriate²³. The 2005 Convention on the Protection and Promotion of the Diversity of Cultural Expressions reaffirmed the cultural-rights dimension of heritage emphasising pluralism and equitable access to cultural expressions, coupled with the affirmation of the sovereign right of States to adopt cultural policies in conformity with their human rights commitments, thus reinforcing the linkage between cultural diversity and cultural rights²⁴. These treaty regimes operate alongside a substantial body of soft-law instruments with significant normative influence.

¹⁷ E. CUNLIFFE, P. FOX, *Safeguarding Cultural Property and the 1954 Hague Convention: All Possible Steps*, Woodbridge, 2022, pp. 3-39.

¹⁸ M. ZIA-UD-DIN, *The International Legal Framework Governing the Protection of Cultural Heritage*, in *FWU Journal of Social Sciences*, vol. 18, 2024, pp. 35-43. See also J. HLADIK, *Reporting system under the 1954 Convention for the Protection of Cultural Property in the Event of Armed Conflict*, in *International Review of the Red Cross*, vol. 82, 2000, pp. 1001-1016; T. DESCH, *The Second Protocol to the 1954 Hague Convention for the Protection of Cultural Property in the Event of Armed Conflict*, in *Yearbook of International Humanitarian Law*, vol. 2, 1999, pp. 63-90; J. D. KILA, *Heritage under siege: military implementation of cultural property protection following the 1954 Hague Convention*, Leiden/Boston, 2012.

¹⁹ R. O'KEEFE, *World cultural heritage: obligations to the international community as a whole?*, in *International & Comparative Law Quarterly*, vol. 53, 2004, pp. 189-209.

²⁰ L. V. PROTT, *International control of illicit movement on the cultural heritage: The 1970 UNESCO convention and some possible alternatives*, in *Syracuse Journal of International Law and Commerce*, 1983, pp. 333-351.

²¹ H. CLEERE, *The concept of "outstanding universal value" in the World Heritage Convention*, in *Conservation and Management of Archaeological Sites*, 1996, pp. 227-233.

²² L. LIXINSKI, *Intangible cultural heritage in international law*, Oxford, 2013.

²³ J. BLAKE, *Engaging "communities, groups and individuals" in the International mechanisms of the 2003 Intangible Heritage Convention*, in *International Journal of Cultural Property*, vol. 26, 2019, pp. 113-137. See also J. BLAKE, UNESCO's 2003 Convention on Intangible Cultural Heritage: The implications of community involvement in 'safeguarding', in L. SMITH, N. AKAGAWA (eds.), *Intangible Heritage*, London, 2008, pp. 45-73.

²⁴ Y. DONDEERS, *Do cultural diversity and human rights make a good match?*, in *International social science journal*, vol. 61, 2010, pp. 15-35.

The Faro Convention of the Council of Europe frames cultural heritage as a substantive right, founded in democratic participation and thus reinforcing the shift towards a right-based approach to heritage governance²⁵. ICOMOS charters and UNESCO Operational Guidelines further specify technical standards relating to authenticity, integrity, community involvement and safeguarding obligations, and have acquired *de facto* normative authority through consistent State practice and their integration into operational monitoring mechanisms²⁶. International and regional jurisprudence provides additional consolidation.

The Inter-American Court and Commission have interpreted cultural heritage within the corpus of indigenous rights, emphasising continuity and collective memory as protected legal interests. The UN Human Rights Committee has read Art. 27 ICCPR as safeguarding minority cultural practices, including traditional land-related activities, while regional courts - including the European Court of Human Rights, which has recognised cultural identity and community life as protected interests within the scope of Art. 8 ECHR - have incorporated cultural dimensions into their jurisprudence²⁷. The ICC's Al Mahdi judgment confirmed that intentional destruction of cultural heritage constitutes a serious breach of international law insofar as it directly affects the identity and collective memory of protected groups²⁸.

The case concerned the deliberate targeting of mausoleums in Timbuktu in 2012 and marked the first time an international criminal tribunal recognised the destruction of cultural heritage as a standalone war crime under Article 8(2)(e)(iv) of the Rome Statute²⁹.

Taken together, this corpus articulates a multilayered body of norms under which cultural heritage, tangible and intangible, is protected both as a collective interest and as an expression of cultural rights. It also provides the substantive and procedural *criteria* against which emerging technologies must be assessed, ensuring that AI-driven heritage practices align with established international obligations on safeguarding and cultural-rights protection, while still revealing areas where existing instruments have yet to fully accommodate the regulatory challenges posed by artificial intelligence.

2.1. Cultural heritage and human rights

The integration of cultural heritage into the field of international human rights has followed a separate normative trajectory from the development of heritage-specific treaty

²⁵ W. LOGAN, Cultural diversity, cultural heritage and human rights: towards heritage management as human rights-based cultural practice, in S. EKERN, W. LOGAN, B. SAUGE, A. SINDING-LARSEN (eds.), *World heritage management and human rights*, London, 2016, pp. 19-32.

²⁶ C. BORTOLOTO, UNESCO, cultural heritage, and outstanding universal value: value-based analyses of the World Heritage and Intangible Cultural Heritage Conventions, in *International Journal of Heritage Studies*, vol. 21, 2015, pp. 528-530.

²⁷ F. FRANCONI, Culture, Heritage and Human Rights: an Introduction, in F. FRANCONI, M. SCHEININ (eds.), *Cultural Human Rights*, Leiden, 2008, pp. 1-15.

²⁸ K. WIERCZYŃSKA, A. JAKUBOWSKI, Individual responsibility for deliberate destruction of cultural heritage: contextualizing the ICC judgment in the Al-Mahdi case, in *Chinese Journal of International Law*, vol. 16, 2017, pp. 695-721.

²⁹ P. ROSSI, The "Al Mahdi" Trial Before the International Criminal Court: Attacks on Cultural Heritage Between War Crimes and Crimes Against Humanity, in *Diritti umani e diritto internazionale*, 2017, pp. 87-100. See also M. M. SADOWSKI, Heritage strikes back: the Al Mahdi case, ICC's policy on cultural heritage and the pushing of law's boundaries, in *Undecidabilities and Law*, 2022, pp. 99-119; I. Salsano, The Absent Framework: A Legal Semiotics Inquiry into the Destruction of Collective Memory through Cultural Heritage Loss, in *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique*, 2025, pp. 1-27.

regimes³⁰. From the 1960s onwards, the consolidation of cultural rights as part of the universal human-rights *corpus* provided a new conceptual basis for understanding heritage as an interest rooted in human dignity. This evolution introduced rights-based protections for cultural identity and participation, grounded in human-rights obligations and operating independently of UNESCO's preservation-oriented regime³¹. Art. 27 ICCPR constitutes the first explicit recognition of cultural entitlements in universal human-rights law³².

By requiring States to ensure that minorities enjoy their own culture, Art. 27 establishes: (i) a substantive guarantee protecting cultural practices as elements of group identity; (ii) a continuity requirement that obliges States to prevent measures that would interrupt the intergenerational transmission of such practices; and (iii) a spatial dimension, whereby cultural rights extend to land-related practices, territories and knowledge systems essential to the maintenance of heritage³³. In parallel, Art. 15 ICESCR introduces a complementary dimension of protection by recognising the right of everyone to take part in cultural life and shifting the focus from minority status to the general conditions under which cultural participation is possible³⁴. Art. 15 imposes positive obligations on States and requires ensuring the availability of cultural institutions and resources, their physical and economic accessibility and the existence of appropriate institutional arrangements for the preservation, documentation and transmission of cultural expressions³⁵.

In this perspective, the material and immaterial components of heritage constitute part of the infrastructure through which the right to participate in cultural life is realised.

Read together, these articles configure a dual axis of protection: on the one hand, group-specific guarantees that shield the cultural practices of minorities from disruption and erasure; on the other, general obligations to create and maintain an enabling environment for cultural participation. Once cultural participation acquired the status of protected legal positions, cultural heritage itself entered the normative orbit of human rights.

A series of interpretive developments consolidated this shift.

First, heritage came to be regarded not merely as an object of preservation but as one of the conditions for the exercise of cultural rights. This entails that States must regulate heritage in a manner that does not impair the material, spatial and institutional environments upon which cultural practices rely³⁶. As clarified in the practice of human-rights bodies, the protection of cultural rights therefore extends to the maintenance of contextual conditions: an approach consistently affirmed by the Human Rights Committee in General Comment

³⁰ A. F. VRDOLJAK, Human Rights and Culture Heritage in International Law, in F. LENZERINI, A. F. VRDOLJAK (eds.), *International Law for Common Goods: Normative Perspectives on Human Rights, Culture and Nature*, Oxford, 2013, pp. 139-174.

³¹ H. SILVERMAN, D. F. RUGGLES, *Cultural Heritage and Human Rights*, New York, 2007. See also L. F. HASHIMOVA, *The Development of the Human Rights Factor in Cultural Heritage Protection: Interrelationships and New International Legal Trends*, in *Vestnik of Saint Petersburg University Law*, 2025, pp. 151-166.

³² S. HOSSEIN, *Article 27 of the Universal Declaration, Preserving Individual Rights or Mechanism for Support Collective Ones*, in *International Journal of Humanities & Social Science Studies (IJHSSS)*, 2017, pp. 203-228.

³³ G. ANDRÁSSY, *The Real Meaning of Article 27 ICCPR and the Deeper Understanding of Overlapping Human Rights Its Exploration Requires*, in *Austrian Review of International and European Law Online*, vol. 28, 2025, pp. 61-93.

³⁴ Y. DONDEERS, *Foundations of collective cultural rights in international human rights law*, in *Amsterdam Law School Research Paper*, 2015, pp. 1-19.

³⁵ R. O'KEEFE, *The "right to take part in cultural life" under article 15 of the ICESCR*, in *International & Comparative Law Quarterly*, vol. 47, 1998, pp. 904-923.

³⁶ J. BLAKE, *Taking a Human Rights Approach to Cultural Heritage Protection*, in *Heritage & Society*, 2011, pp. 199-238.

No. 23³⁷, which interprets Art. 27 as protecting the territorial, environmental and practical circumstances enabling cultural practices. The Committee on Economic, Social and Cultural Rights in General Comment No. 21 identifies institutional, linguistic and knowledge-transmission structures as essential components of the effective enjoyment of Art. 15, without which the right to take part in cultural life cannot be effectively realised³⁸.

Human-rights bodies have emphasised that States must safeguard: (i) access to the material and immaterial resources on which protected practices depend - including ceremonial sites or ritual artefacts - as repeatedly affirmed by the Human Rights Committee under Art. 27; (ii) the spatial and functional characteristics of culturally significant areas, whose alteration may compromise the exercise of cultural practices, a point underscored by the Inter-American Court of Human Rights in cases such as *Yakye Axa v. Paraguay*³⁹ and *Kaliña and Lokono Peoples v. Suriname*⁴⁰; and (iii) intergenerational transmission mechanisms, identified as prerequisites for the effective enjoyment of Art. 15.

Secondly, the human-rights framework has progressively limited the discretion traditionally associated with heritage governance. Measures affecting heritage now fall within the scope of rights-based scrutiny, requiring States to demonstrate that any restriction on cultural practices or access to cultural resources pursues a legitimate aim, follows due process and satisfies tests of necessity and proportionality. This evolutionary development shows how cultural heritage is governed through legal standards that generate reviewable obligations. Finally, the emergence of a rights-holder-centred perspective has reconfigured the position of communities within heritage regimes: they now hold legally protected interests relating to identity and transmission⁴¹. This includes the right to be consulted, to influence decisions affecting their heritage, and to contest measures that materially alter or decontextualise the cultural resources on which their practices depend.

³⁷ General Comment No. 23 of the Human Rights Committee of 8 April 1994, *Article 27 (Rights of Minorities)*, CCPR/C/21/Rev.1/Add.5, p. 3.

³⁸ General Comment No. 21 of the Committee on Economic, Social and Cultural Rights of 21 December 2009, *Right of everyone to take part in cultural life (Article 15, para. 1 (a))*, E/C.12/GC/21, p. 12.

³⁹ IACtHR - Inter-American Court of Human Rights, *Case of the Yakye Axa Indigenous Community v. Paraguay*. Judgment of June 17, 2005, Merits, Reparations and Costs, p. 81. The case concerned the longstanding inability of the *Yakye Axa* Community to regain access to its ancestral lands, following their transfer to private ownership and the State's failure to ensure restitution or effective territorial protection. As a result, the community was forced to reside in a precarious roadside settlement without access to the resources and spaces necessary for its cultural and subsistence practices. In para. 50.15, the Court observed that the relocation to *Estancia El Estribo* did not remedy these conditions and, critically, prevented the community from freely conducting its cultural practices. It held that the new settlement lacked the environmental, territorial and resource conditions required for cultural continuity, thereby affirming that the protection of cultural integrity is inseparable from securing effective access to traditional lands and the material context enabling the exercise of protected cultural practices. See also K. ZOMBORY, *The Right to Cultural Identity in the Case Law of the Inter-American Court of Human Rights: A New Global Standard for the Protection of Indigenous Rights and Future Generations?* in *Journal of Agricultural and Environmental Law*, vol. 18, 2023, pp. 171-191.

⁴⁰ IACtHR - Inter-American Court of Human Rights, *Case of the Kaliña and Lokono Peoples v. Suriname*. Judgment of November 25, 2015 (Merits, Reparations and Costs), p. 48. The case concerned the longstanding failure of Suriname to recognise and protect the traditional territories of the *Kaliña* and *Lokono* peoples, including areas of ceremonial, spiritual and subsistence significance. The Court noted that the concession of mining, logging and nature-reserve projects within indigenous lands had been authorised without adequate consultation and had resulted in the degradation of sites essential to cultural identity and transmission. See also L. LIXINSKI, *Case of the Kaliña and Lokono Peoples v. Suriname*, in *American Journal of International Law*, vol 111, 2017, pp. 147-154.

⁴¹ L. LIXINSKI, *International Heritage Law for Communities: Exclusion and Re-imagination*, Oxford, 2019.

These developments have expanded the reach of cultural rights beyond the protection of individual expressions to encompass the broader normative environment in which heritage is produced, maintained and transmitted.

3. Cultural heritage and human rights in the digital age

In the digital age, the intersection between cultural heritage, human rights and emerging technologies has acquired a greater complexity. Key issues as the essence of cultural rights, preservation of collective memory, and utilisation of AI to safeguard the multicultural heritage of mankind emerge as the technological revolution refigures the way we interact with heritage. Digitisation and AI are both an opportunity and a source of tension for heritage protection, particularly in relation to data governance and the risk of reproducing dominant epistemologies in digital cultural ecosystems⁴².

The digital transformation generates unprecedented benefits for accessibility and conservation but simultaneously intensifies existing human-rights concerns⁴³.

Cultural heritage is the physical expression of our shared human past as tangible artefacts, intangible traditions, languages, rituals, and knowledge passed down through generations⁴⁴. UNESCO provides the legal framework for understanding cultural heritage in its twofold nature: according to the 2003 UNESCO Convention heritage encompasses «the practices, representations, expressions, knowledge, skills [...] that communities, groups and, in some cases, individuals recognise as part of their cultural heritage»⁴⁵.

This definition confirms that heritage extends beyond monuments and artefacts to encompass linguistic practices, rituals, oral traditions. In doing so, the 2003 Convention broadened the focus of earlier UNESCO instruments - most notably the 1954 Hague Convention and the 1972 World Heritage Convention⁴⁶ - which were primarily centred on the protection of tangible cultural property and monumental heritage.

Given this shift, intangible cultural heritage has become a central field of legal and digital intervention, after the 2003 Convention elevated the safeguarding of traditional

⁴² F. A. RASO, H. HILLIGOSS, V. KRISHNAMURTHY, C. BAVITZ, L. KIM, *Artificial intelligence & human rights: Opportunities & risks*, in *Berkman Klein Center Research Publication*, 2018.

⁴³ E. RISSE, *Human Rights and Artificial Intelligence: an urgently needed Agenda*, in *Human Rights Quarterly*, 2019, pp. 1-16.

⁴⁴ I. SALSANO, *La tutela giuridica internazionale del patrimonio culturale intangibile dei migranti*, in *Eunomia. Rivista di studi su pace e diritti umani*, vol. 13, 2024, pp. 187-212.

⁴⁵ J. BLAKE, *Safeguarding Intangible Cultural Heritage*, in F. FRANCIONI, A. F. VRDOLJAK (eds.), *The Oxford Handbook of International Cultural Heritage Law*, Oxford, 2020, pp. 347-370.

⁴⁶ In addition, early UNESCO and UN-system instruments adopted between the 1960s and 1990s consolidated a predominantly State-centric approach to heritage. Resolutions of the UNESCO General Conference (e.g., 1964 Recommendation on the Safeguarding of Beauty and Character of Landscapes and Sites, 1976 Recommendation on Participation of the Public in Cultural Life) framed heritage protection as a matter of administrative conservation rather than rights-based entitlement. Similarly, the 1982 World Conference on Cultural Policies advanced a developmental reading of culture as a factor of national identity and economic progress but did not yet articulate the community-centred or rights-holder logic later embraced by the 2003 Convention. These instruments collectively illustrate the pre-2003 paradigm: heritage conceived as an object of management by States, rather than as a domain of legally protected community agency.

knowledge and oral practices to an international obligation⁴⁷. The boundary between tangible and intangible heritage has, however, become increasingly opaque in the digital age: heritage arises from the interaction between communities and their environments, producing manifestations ranging from languages and rituals to buildings and movable objects⁴⁸.

The 2003 Convention expressly acknowledged this interdependence, and later the Faro Convention further repositioned people at the normative centre of heritage governance⁴⁹.

The spirit of a place, recognised by the International Council on Monuments and Sites⁵⁰, includes the tangible and intangible elements that constitute the values related to a specific *locus*. This holistic understanding of cultural heritage is essential for addressing contemporary challenges⁵¹. Cultural heritage functions as both a material and immaterial repository of collective memory and identity; its protection therefore extends beyond preservation *stricto sensu* and engages substantive human-rights obligations.

In this perspective, safeguarding heritage constitutes a prerequisite for the effective enjoyment of cultural rights, which explains the progressive recognition of cultural heritage interests within the *corpus* of international human-rights law. A glance at the International Bill of Rights, as shown in par. 2.1., confirms this trajectory, as both ICCPR and ICESCR have progressively been interpreted in a manner that incorporates heritage-related interests into the scope of cultural rights⁵². Within international human-rights law, the full realisation of cultural rights requires a tripartite structure encompassing access to cultural resources, participation in cultural life, and the enjoyment of cultural expressions⁵³.

A core component of this triad is intergenerational cultural memory, through which communities transmit, reinterpret and preserve their histories⁵⁴. Its protection is essential for the exercise of cultural rights and for the maintenance of collective identity in contemporary societies⁵⁵. Cultural memory, however, is increasingly shaped by digital processes.

What once relied on material repositories and community-based transmission is now mediated by digital infrastructures determining how cultural narratives are stored and accessed⁵⁶. Digital technologies, initially employed for documentation, have evolved into systems that influence the transmission and interpretation of collective memory⁵⁷.

⁴⁷ Y. HOU, S. KENDERDINE, D. PICCA, M. EGLOFF, A. ADAMOU, *Digitizing intangible cultural heritage embodied: State of the art*, in *Journal on Computing and Cultural Heritage*, 2022, pp. 1-20.

⁴⁸ D. H. R. SPENNEMANN, *ChatGPT and the generation of digitally born knowledge: How does a generative AI language model interpret cultural heritage values?*, in *Knowledge*, 2023, pp. 480-512.

⁴⁹ B. ROUHANI, *Ethically digital: contested cultural heritage in digital context*, in *Studies in Digital Heritage*, 2023, pp. 1-16.

⁵⁰ ICOMOS from now on.

⁵¹ S. PANSONI, S. TIRIBELLI, M. PAOLANTI, F. DI STEFANO, E. FRONTONI, E. S. MALINVERNI, B. GIOVANOLA, *Artificial intelligence and cultural heritage: design and assessment of an ethical framework*, in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2023, pp. 1149-1156.

⁵² S. EKERN, W. LOGAN, B. SAUGE, A. SINDING-LARSEN, *Human rights and World Heritage: preserving our common dignity through rights-based approaches to site management*, in *International Journal of Heritage Studies*, 2012, pp. 213-225.

⁵³ F. LENZERINI, *Intangible Cultural Heritage: The Living Culture of Peoples*, in *European Journal of International Law*, 2011, pp. 101-120.

⁵⁴ G. NEGLIA, M. ANGRISANO, I. MECCA, F. FABBROCINO, *Cultural heritage at risk in world conflicts: Digital tools contribution to its preservation*, in *Heritage*, 2024, pp. 6343- 6365.

⁵⁵ Y. DONDEERS, *Cultural heritage and human rights*, in F. FRANCONI, A. F. VRDOLJAK (eds.), *The Oxford Handbook on International Cultural Heritage Law*, Oxford, 2020, pp. 379-406.

⁵⁶ N. MOHAMAD BA'AI, A. ARIS, *AI and cultural heritage: Preserving and promoting global cultures through technology*, in *Nanotechnology Perceptions*, 2024.

⁵⁷ I. SILIUTINA, ET AL., *Cultural preservation and digital heritage: challenges and opportunities*, in *Amazonia Investiga*, 2024, pp. 262-273; see also F. HASSANI, *Documentation of cultural heritage: techniques, potentials, and constraints*, in *The*

The rapid development of digital technologies has expanded their function beyond documentation and conservation, integrating them into interpretive processes and heritage management. These tools enable the structured collection and verification of cultural data, support integrity controls, and provide advanced scanning and modelling capacities essential for risk mitigation and restoration. Such technical infrastructures create unprecedented possibilities for preserving and transmitting cultural memory in digital form, across both geographical and temporal boundaries, thereby reshaping how heritage is accessed and managed⁵⁸. These developments have also led to the emergence of large digital platforms acting as governance infrastructures for cultural data. Platforms such as *Europeana*, which allows users to explore European cultural heritage through online browsing of more than 58 million artefacts⁵⁹, the UNESCO World Heritage List Portal⁶⁰, or similarly, the Time Machine project⁶¹, are a step forward in making cultural heritage available to a large audience of stakeholders⁶². Their role is particularly relevant for diasporic or disabled groups, for whom digital access may constitute the primary or sole means of reconnecting with cultural resources. Advanced digitisation techniques, including 3D scanning, photogrammetry, augmented-reality interfaces and AI-assisted restoration tools, enhance the capacity to document and safeguard endangered cultural materials⁶³. AI systems expand accessibility by enabling multilingual search, automated description, and adaptive interfaces, thereby broadening participation in cultural life and facilitating exposure to diverse historical narratives⁶⁴. From a legal perspective, these platforms also operate as normative infrastructures: their metadata schemes influence how cultural works are classified and circulated. While digital infrastructures significantly expand the potential for access to cultural heritage, their democratising promise remains unevenly realised⁶⁵.

Large-scale digitisation and AI initiatives are often designed and funded by Global North institutions, which exercise control over technical standards, metadata schemas and classificatory logics. Source communities, particularly in the Global South, may have limited influence over digitisation priorities, contextual information or conditions of reuse⁶⁶.

International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2015, pp. 207-214; D. H. HAUX, A. MAGET DOMINICÉ, J. A. RASPOINIG, *A cultural memory of the digital age?*, in *International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique*, 2021, pp. 769-782.

⁵⁸ S. ALLEGREZZA, *L'intelligenza artificiale applicata alla digitalizzazione del patrimonio culturale*, in *Il cinquantenario della Convenzione UNESCO a tutela del patrimonio culturale (1972-2022): percorsi multidisciplinari*, 2022, pp. 9-16.

⁵⁹ B. CARAMIAUX, *The use of artificial intelligence in the cultural and creative sectors – Concomitant expertise for INI report*, in *Research for CULT Committee*, 2020.

⁶⁰ A. É. GFELLER, J. EISENBERG, UNESCO and the shaping of global heritage, in P. DUEDAHL (eds.), *A history of UNESCO: Global actions and impacts*, London, 2016, pp. 279-299.

⁶¹ A European research initiative that develops computational models and AI-assisted tools to reconstruct, map and analyse historical data across centuries. By integrating archives, museum collections and geospatial datasets into interoperable digital infrastructures, it enables long-term preservation and cross-disciplinary access to cultural heritage resources.

⁶² I. DI LENARDO, F. KAPLAN, *Venice Time Machine: Recreating the density of the past*, in *Digital Humanities 2015*, 2015.

⁶³ A. M. SULLIVAN, *Cultural heritage & new media: a future for the past*, in *J. Marshall Review of Intellectual Property Law*, 2015, pp. 605-645.

⁶⁴ P. K. YU, *Cultural relics, intellectual property, and intangible heritage*, in *Temple Law Review*, 2008, pp. 434-506.

⁶⁵ L. IACOVINO, *Shaping and reshaping cultural identity and memory: maximising human rights through a participatory archive*, in *Archives and Manuscripts*, 2015, pp. 29-41.

⁶⁶ L. LIXINSKI, *Digital heritage surrogates, decolonization, and international law: restitution, control, and the creation of value as reparations and emancipation*, in *Santander Art and Culture Law Review*, 2020, 65-86.

This produces structural asymmetries in authorship and benefit-sharing, as collections are digitised and integrated into transnational platforms without adequate consultation or control over downstream uses⁶⁷. Against this background, these questions arise: who controls the metadata and taxonomies by which cultural materials are classified? Who determines which sources are treated as authoritative for contextualisation?

To what extent are cultural expressions protected when they are removed from their social environment and recast as digital assets or training data?⁶⁸

These asymmetries have prompted a body of soft-law instruments specifically addressing the legal implications of AI and digitisation for cultural-heritage governance.

At the international level, the Council of Europe's "Guidelines given the latest technological developments, such as AI, complementing Council of Europe standards in the fields of culture, creativity and cultural heritage"⁶⁹ constitute the first attempt to translate existing CoE cultural-rights standards into operational requirements for the design, deployment and oversight of AI systems in the cultural sector⁷⁰.

The Guidelines explicitly complement the Council of Europe's *acquis* on cultural diversity and democratic participation, requiring member States to: (i) ensure transparency for algorithmic tools used for heritage classification, curation or access; (ii) conduct *ex ante* and ongoing risk assessments addressing algorithmic distortion; (iii) secure effective community participation in decisions concerning metadata; and (iv) establish accountability and redress mechanisms, including traceability obligations, provenance documentation against the decontextualised use of cultural datasets in training AI models.

The Guidelines frame digital heritage as a domain that requires rights-based regulatory oversight rather than purely technical management⁷¹. At the EU level, the European Parliament's Resolution of 19 May 2021 on Artificial Intelligence in education, culture and the audiovisual sector sets out sector-specific normative expectations for AI systems used in cultural contexts, stressing that they must respect the Union's commitment to cultural and linguistic diversity under Artt. 2 and 3 TEU and Art. 167 TFEU and highlights the risk that algorithmic systems and content-ranking mechanisms may amplify dominant cultural products while marginalising minority expressions⁷².

The Resolution therefore calls for human oversight on cultural-sector use cases, as well as safeguards against exclusionary practices in automated curation.

It also underlines the responsibility of platforms and cultural intermediaries to ensure that AI-enabled tools do not undermine the diversity of cultural expressions or the rights of authors and communities⁷³. The EU Commission Recommendation 2021/1970 on a

⁶⁷ R. GRIFFIN, *Rethinking rights in social media governance: human rights, ideology and inequality*, in *European Law Open*, 2023, pp. 30-56.

⁶⁸ D. GIGLITTO, L. CIOLFI, C. CLAISSE, E. LOCKLEY, *Bridging cultural heritage and communities through digital technologies: understanding perspectives and challenges*, in *C&T 2019 - Proceedings of the 9th International Conference on Communities and Technologies: Transforming Communities*, New York, 2019.

⁶⁹ Adopted on 13 December 2024 by the Steering Committee for Culture, Heritage and Landscape.

⁷⁰ E. C. DIAZ GALAN, *The Legal Regulation of Artificial Intelligence (AI) in Europe: Two Decisive (but Insufficient) Steps of the Council of Europe and the European Union*, in *Cuadernos Derecho Transnacional*, vol. 17, 2025, p. 366.

⁷¹ C. NARDOCCI, *Artificial Intelligence at the Crossroads between the European Union & the Council of Europe: Who Safeguards What & How?*, in *Italian Journal of Public Law*, vol. 16, 2024, p. 165.

⁷² E. H. MORAWSKA, *Council of Europe standards and activities related to AI: towards a Framework Convention on AI and human rights?*, in M. BALCERZAK, J. KAPELASKA-PREGOWSKA (eds.) *Artificial Intelligence and International Human Rights Law*, Cheltenham, 2024, pp. 25-44.

⁷³ F. P. LEVANTINO, F. PAOLUCCI, *Advancing the Protection of Fundamental Rights Through AI Regulation: How the EU and the Council of Europe are Shaping the Future*, in *European Yearbook on Human Rights 2024*, 2024, pp. 3-37.

common European data space for cultural heritage establishes a governance framework for the lawful processing, interoperability and reuse of digitised cultural resources.

It invites member States and cultural heritage institutions to develop data infrastructures that facilitate access and reuse while complying with data-protection, intellectual-property and cultural-heritage norms⁷⁴. The Recommendation identifies risks arising from the deployment of AI models trained on cultural datasets, such as the loss of provenance information and potential misappropriation of community-specific materials.

To address these risks, it urges States to adopt mechanisms ensuring provenance integrity and rights-compliant data-sharing practices and stresses the need for open and verifiable data architectures capable of supporting scrutiny and community participation in the digital heritage management. These instruments shape the environment within which States must reconcile AI-enabled heritage practices with their obligations under international human-rights law, articulating a set of culture-specific normative parameters for algorithmic governance and regulatory approaches grounded in cultural rights and in the protection of cultural memory.

4. *Ethical dilemmas and normative frictions in AI-driven cultural heritage practices*

While AI becomes increasingly integrated into heritage practices through digitisation, interpretation and simulation, profound normative conflicts related to authorship and authenticity are compounded⁷⁵. AI reconstructions, for instance, realised with generative algorithms and neural rendering, can now virtually bring to life destroyed ancient monuments⁷⁶. Such efforts often claim to restore memory, but they raise critical legal and ethical issues⁷⁷. Who owns the reconstructed model? Is it a neutral act of remembrance or a new form of appropriation? Who decides which version of the past is digitally recreated?

The issue of digital authenticity emerges forcefully here.

When AI reconstructs missing architectural elements or produces art representations, it does so through probabilistic inference, training data, and algorithmic pattern recognition.

The output is not a genuine *facsimile* of lost heritage but a machine-authored assumption, based on developer choices and institutional agendas, revealing how the appearance of objectivity is often illusory as algorithms encode value judgments, assumptions and exclusions⁷⁸. When AI is deployed in metadata generation, artefact classification or interpretive interfaces such as chatbots and interactive museum guides, its outputs reflect the

⁷⁴ P. FORADORI, P. ROSA, *Guardian of culture: the European Union's quest for actorness in the protection of cultural heritage in conflicts and crises*, in *European Security*, 2025, pp. 1-22.

⁷⁵ S. COLLEY, *Ethics and digital heritage*, in T. IRELAND, J. SCHOFIELD (eds.), *The ethics of cultural heritage*, New York, 2014, pp. 13-32.

⁷⁶ E. L. THOMPSON, *Legal and ethical considerations for digital recreations of cultural heritage*, in *Chapman Law Review*, 2017, pp. 153-176. See also Z. MANŽUCH, *Ethical issues in digitization of cultural heritage* in *Journal of Contemporary Archival Studies*, 2017, pp. 1-17.

⁷⁷ B. ROUHANI, *Ethically Digital: Contested Cultural Heritage in Digital Context*, in *Studies in Digital Heritage*, vol. 7, 2023, pp. 1-16.

⁷⁸ G. GUARINO, *Intelligenza artificiale generativa e memoria storica: la problematica del revisionismo tra condiscendenza e allucinazioni*, in *Futuri*, 2025, pp. 251-264.

biases embedded in the training data, that privilege hegemonic cultural models - Western art canons, colonial classificatory systems - and flatten minor histories⁷⁹.

Moreover, AI systems often lack transparency and when they misclassify a sacred object or misinterpret a traditional practice, affected communities have little recourse, legally or symbolically, even though technical correction alone is not sufficient⁸⁰ and their human input is crucial⁸¹. As AI becomes more embedded in heritage contexts, ethical adequacy depends less on technical accuracy and more on interpretive empathy and cultural sensitivity: human beings possess a depth of contextual and emotional interpretation that no algorithm can provide, namely, the ability to value the symbolic, spiritual and social forms of heritage⁸².

Digitisation also exposes enduring power imbalances: many collections originate from colonial seizure or unconsented acquisition, and their cataloguing and digital representation should reflect historical context and source community perspectives⁸³.

Cultural heritage is a living set of historical, spiritual and communal relationships and when AI systems simulate cultural expressions without the free, prior, and informed consent of the concerned communities, they risk enabling a new form of digital dispossession⁸⁴, especially in cases of indigenous knowledge⁸⁵. Treaties like the UNESCO Convention on intangible heritage recognise the supremacy of community participation but are poorly situated to regulate algorithmic intermediation. Declarations like the UN Declaration on the Rights of Indigenous Peoples enshrine indigenous peoples' rights to their cultural manifestations but provide no lead on the implications of AI-automated curations⁸⁶.

The 2021 UNESCO Recommendation on the Ethics of AI is non-binding and does not engage with culture-specific algorithmic risks; likewise, the EU AI Act and the EU framework on trustworthy AI contain no provisions targeting heritage-related harms⁸⁷.

This normative gap leaves cultural heritage vulnerable: increasingly digitised and AI-mediated yet insufficiently protected by legal and ethical safeguards. Communities are often excluded from the design and deployment of AI systems that reshape their cultural presence online. If cultural heritage is to sustain continuity and identity, AI must operate within a framework that treats cultural rights as non-negotiable, and this requires adapting existing instruments to the digital sphere and encouraging communities' participation in governance

⁷⁹ G. MANCINI PALAMONI, *Lo sviluppo sostenibile del patrimonio culturale tra emergenze e tecnologie digitali*, in *Rivista italiana di informatica e diritto*, 2022, pp. 1-12.

⁸⁰ A. FOKA, G. GRIFFIN, D. O. PABLO, P. RAJKOWSKA, and S. BADRI, *Tracing the bias loop: AI, cultural heritage and bias-mitigating in practice*, in *AI & SOCIETY*, 2025, pp. 1-13.

⁸¹ K. GHATH, *AI integration in cultural heritage conservation-Ethical considerations and the human imperative*, in *International Journal of Emerging and Disruptive Innovation in Education*, 2024, pp. 1-10.

⁸² P. AHRWEILER, *Towards Culture-Sensitive, Responsive, and Participatory AI*, in P. AHRWEILER (eds.), *Participatory Artificial Intelligence in Public Social Services: From Bias to Fairness in Assessing Beneficiaries*, Cham, 2025, pp. 277-306.

⁸³ K. PRAZMOWSKA-MARCINOWSKA, *Misappropriation of Indigenous cultural heritage - intellectual property rights in the digital era*, in *Santander Art and Culture Law Review*, 2020, pp. 119-150.

⁸⁴ B. J. MARY, *Artificial Intelligence, Cultural Heritage, and International Relations: The Ethical and Political Dimensions of Digital Conservation*, University of Kingston.

⁸⁵ R. BALLARDINI, I. TUOMINEN, and D. GIRARDI, *Digital Indigenous cultural heritage - Reconciling intellectual property rights, human rights and Indigenous worldviews through proactive approaches and engagement*, in MD. J. HOSSAIN BHUIYAN, M. RAFIQU ISLAM (eds.), *Business, human rights and sustainable development*, Leiden/Boston, 2024, pp. 406-431.

⁸⁶ D. CAMBOU, *The UNDRIP and the legal significance of the right of indigenous peoples to self-determination: a human rights approach with a multidimensional perspective*, in D. SHORT, C. LENNOX, J. BURGER, J. HOHMANN (eds.), *The United Nations Declaration on the Rights of Indigenous Peoples*, London, 2020.

⁸⁷ J. LAUX, *Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act*, in *AI & society*, 2024, pp. 2853-2866.

models. The ethical issues raised by AI - ownership, representation, cultural autonomy - expose the limits of current legal tools and reveal a broader normative deficit in international cultural heritage law: the emergence of digital dispossession as a new form of cultural erosion, which exposes the limits of existing human-rights protections and underscores the need to recognise a collective right to memory in the digital environment.

In this light, the digital transformation of cultural heritage must be approached not merely as a technical advancement but as a normative evolution that reshapes the epistemic foundations of heritage preservation and representation⁸⁸.

Choices in training data and metadata may reinforce historical imbalances and epistemic asymmetries unless subject to clear regulatory constraints.

The absence of apt legal standards raises fundamental questions: who defines the meaning of algorithmically mediated heritage, how can distortions or exclusions be challenged, and what forms of redress are available when cultural histories are altered by digital infrastructures? Given the current lack of authoritative guidance, AI-driven heritage practices require a human-rights-informed framework to ensure their legal and ethical legitimacy⁸⁹ against epistemic displacement⁹⁰. This invites a broader rethinking of legal protections for cultural communities in the digital age: one that recognises not only the materiality of heritage, but its symbolic, participatory, and collective mnemonic dimensions as an emerging legal imperative⁹¹.

5. *From cultural erosion to legal recognition: towards a right to collective memory in digital heritage governance*

As AI increasingly mediates cultural heritage processes, whether through digitisation, classification, or generative reconstruction, it not only reorganises memory access but reshapes the very ontological structures that sustain it.

What emerges is a juridically significant transformation: cultural memory, long protected as a diffuse interest under cultural rights regimes, is now structurally exposed to algorithmic distortion and institutional asymmetries. The ongoing digital transformation of heritage governance is reshaping the relationship between communities, their memories and the mechanisms through which cultural meaning is preserved and transmitted.

Large AI models operate as memory agents, absorbing vast bodies of cultural knowledge, while technology companies increasingly position themselves as primary gateways to information across widely used platforms⁹². This process can be understood as “digital dispossession”, marked by the erosion of community agency over their cultural

⁸⁸ K. CHAINOGLOU, S. KATSIOS, Threats to Cultural Heritage: Normative Developments on AI and Cultural Heritage, in N. KOUTRAS, N. SELVADURAI (eds.), *Recreating Creativity, Reinventing Inventiveness*, London, 2024.

⁸⁹ L. M. RAFANELLI, *Justice, injustice, and artificial intelligence: Lessons from political theory and philosophy*, in *Big Data & Society*, vol. 9, 2022.

⁹⁰ A. P. STEFANIJA, Power asymmetries, epistemic imbalances and barriers to knowledge: the (im) possibility of knowing algorithms, in S. LINDGREN (eds.), *Handbook of critical studies of artificial intelligence*, Cheltenham, 2023, pp. 563-572.

⁹¹ L. CONTI, Understanding Culture, Cultural Identity, and Cultural Heritage in the Post-Digital Age, in I. MUENSTERMANN (eds.), *Belonging in Culturally Diverse Societies-Official Structures and Personal Customs: Official Structures and Personal Customs*, London, 2025, pp. 65-76.

⁹² J. SCHUH, *AI As Artificial Memory: A Global Reconfiguration of Our Collective Memory Practices?*, in *Memory Studies Review*, 2024, pp. 231-255.

narratives, given that algorithmic mediation determines what gets preserved and how it gets interpreted⁹³. Unlike traditional forms of cultural expropriation, which have historically involved the displacement of material artifacts, digital dispossession operates through subtler, data-driven logics: cultural expressions are extracted, standardised, and recontextualised by AI systems whose design and training often exclude the voices of the very communities to whom such heritage belongs. Memory becomes datasets, traditions are codified into algorithms and historical narrative is managed by infrastructures aligned with institutional agendas⁹⁴. The result is a displacement of representational sovereignty, as communities lose the capacity to shape, contest or even recognise the digital representations affecting them. Existing international instruments are insufficient to address these dynamics.

Art. 15 ICESCR recognises the right to take part in cultural life, and CESCR General Comment No. 21 emphasises identity and participation, but both were formulated in a pre-digital context. They do not regulate algorithmic mediation and provide no specific safeguards against the appropriation of intangible cultural data or the distortion generated by biased training *corpora*. Similar limitations arise in instruments with broader cultural scope.

The UNESCO Convention on the Protection and Promotion of the Diversity of Cultural Expressions focuses primarily on cultural policy and does not regulate the digital reinterpretation of cultural expressions. Overall, these frameworks remain largely silent on issues such as non-consensual digitisation, generative simulations and other AI-driven processes that may alter cultural narratives. The legal foundations for protecting cultural memory can be identified within existing international instruments, but these require substantial development to address contemporary challenges, particularly as the social dimension of cultural heritage has gained prominence in international law.

In this normative gap, the notion of a right to collective memory emerges as a necessary extension of cultural human rights: it should be recognised as a substantive legal interest that communities are entitled to exercise and defend⁹⁵. Collective memory is the communal matrix through which groups build meaning, maintain identity and pass on values from one generation to the next⁹⁶ and cultural heritage grants its preservation because it shapes community's shared past⁹⁷. Therefore, cultural memory is not an ancillary interest but a condition for human dignity and self-determination, requiring protection beyond individual rights frameworks. Such protection must include the participatory conditions under which memory is produced and transmitted⁹⁸, recognising it as a dynamic and inherently collective process⁹⁹. In virtual spaces, this procedure is increasingly mediated by AI infrastructures whose processes are not neutral: settling what is visible, whose past is archived, and how

⁹³ R. PEETERS, M. SCHUILEBURG, The algorithmic society: An introduction, in R. PEETERS, M. SCHUILEBURG (eds.), *The algorithmic society*, London, 2020, pp. 1-15.

⁹⁴ S. MERRIL, Hybrid Methodologies for Studying Social and Cultural Memory in the Postdigital Age, in Q. WANG, A. HOSKINS (eds.), *The Remaking of Memory in the Age of the Internet and Social Media*, Oxford, 2024, pp. 241-257.

⁹⁵ M. HALBWACHS, *On Collective Memory*, Chicago, 1992.

⁹⁶ P. NORA, *Between Memory and History: Les Lieux de Mémoire*, in *Representations*, vol. 26, 1989, pp. 7-24.

⁹⁷ J. V. WERTSCH, H. L. ROEDIGER, *Collective memory: Conceptual foundations and theoretical approaches*, in *Memory*, 2008, pp. 318-326.

⁹⁸ P. Y. S. CHOW, *Culture as collective memories: an emerging concept in international law and discourse on cultural rights*, in *Human Rights Law Review*, 2014, pp. 611-646.

⁹⁹ V. APAYDIN, *Critical Perspectives on Cultural Memory and Heritage: Construction, Transformation and Destruction*, London, 2020.

cultural sense is encoded¹⁰⁰. This normative filtering of memory risks displacing communities from the very narratives that define their identity, as algorithmic logics privilege certain representations while obscuring others. In response to this profound asymmetry of interpretive power, a human right to collective memory would answer to this challenge by suggesting a model under which groups of people have decision-making power over the digital representation and circulation of their cultural stories.

Such a right would encompass access to digital heritage and co-determination over its creation, interpretation and governance, requiring participatory design of AI heritage systems, clear data-governance arrangements and remedies for cultural misrepresentation or algorithmic erasure. Conceiving cultural memory as a human right requires a concrete legal articulation capable of addressing the challenges posed by AI-driven heritage governance, thus rendering a collective memory right essential for ensuring accountability and justice under algorithmic mediation¹⁰¹. Such a right would not require the creation of a new normative order *ex nihilo*. It would instead extend the logic of current cultural-rights frameworks to address the epistemological challenges introduced by digital infrastructures, while aligning with emerging debates on digital rights.

Recognising a right to collective memory would thus enrich the current human-rights architecture by bridging the gap between the cultural specificity of heritage and the structural transformations driven by AI technologies. This right must be operationalised through concrete legal norms: transparency obligations in algorithmic design, consent and co-determination rights, and institutional provisions to guarantee cultural representational sovereignty. This shift requires heritage governance to be reconsidered as an arena of democratic participation, recognising a right to collective memory as necessary to ensure that digital heritage does not reduce to an impersonal archive but remains a living domain of meaning, identity and continuity.

6. Conclusions

The intersection of AI and heritage management requires a recalibration of the normative frameworks regulating culture protection and human rights.

Digital technologies have reshaped the ways in which cultural memory is produced and accessed: heritage, once materially situated and rooted in a community, is now increasingly curated by algorithms that risk reproducing historic asymmetries and sideline minority voices.

This evolution towards digital governance has exposed significant normative frictions between preservation and reappropriation and between community participation and technological mediation. When cultural narratives are recontextualised without the participation or consent of the communities to which they belong, heritage ceases to function as a vector of identity and becomes instead a dataset, legible only through algorithmic logics.

International human rights law, while recognising the right to participate in cultural life and the importance of intangible heritage, remains largely silent on the algorithmic challenges posed by AI, creating a normative gap between the growing influence of AI over cultural

¹⁰⁰ M. S. DE CLIPPELE, *Does the Law Determine What Heritage to Remember?*, in *International Journal for the Semiotics of Law*, vol. 34, 2021, pp. 623-656.

¹⁰¹ D. GUANGYU, *Cultural Heritage Rights and Rights Related to Cultural Heritage: A Review of the Cultural Heritage Rights System*, in *Santander Art and Culture Law Review*, 2023, pp. 167-190.

heritage and the capacity of international law to protect the memory and self-representation of cultural groups. To address this gap, the article proposes a human right to collective memory, grounded in the architecture of cultural human rights yet adapted to the participatory and epistemic needs of the digital era. Such a right would not depart from existing legal traditions but extend them, giving normative recognition to community agency in algorithmic mediation.



ALESSIA DEL PIZZO*

SHAPING FAIR AI: OVERCOMING ALGORITHMIC BIAS AND FOSTERING INCLUSION

SUMMARY: 1. Artificial intelligence as *locus* of conflict between efficiency and justice. – 2. *Bias* and vulnerability: automated discrimination in critical domains. – 3. The role of law: from formal equality to algorithmic inclusion. – 4. Conclusions.

1. *Artificial intelligence as locus of conflict between efficiency and justice*

At the heart of contemporary constitutional democracies lurks a profound tension between the ideal of efficiency – exalted by algorithmic logic¹ – and the principle of substantive justice, the cornerstone of the modern legal system². Artificial intelligence (AI), in its progressive penetration into vital sectors such as healthcare, criminal justice and education, does not merely act as a technical aid, but tends to transform itself into a silent regulatory actor, capable of modulating rights and redefining the relationships between individuals and institutions.

The emergence of AI as an enabling technology marks, in fact, a dislocation of the decision-making function: from the human subject – historically invested with legal responsibility, moral awareness and deliberative capacity – to automated systems that operate according to statistical-probabilistic logic. The concept of “*enabling technology*” should be

* PH.D in Law and Cognitive Neuroscience, Niccolò Cusano University, Rome, Italy.

¹ Algorithmic logic is a set of formal and computational rules that govern the transformation of *input* into *output* by means of a finite sequence of deterministic or probabilistic operations, encoded in an algorithm, and oriented towards the maximisation of efficiency, internal consistency and repeatability of decisions. It replaces or complements human forms of deliberation with automatic mechanisms based on data, models and objective functions. D. E. KNUTH, *The Art of Computer Programming*, Boston, 1997.

² The tension between the ideal of efficiency and substantive justice is a central issue in the contemporary context, in which legal and social institutions are called upon to reconcile economic objectives with ethical values. Efficiency, often understood as the optimisation of time and costs, sometimes runs the risk of clashing with the demands of fairness and justice. While it pursues the quick and pragmatic resolution of conflicts, justice requires a deeper analysis of the circumstances and consequences of decisions. For a more in-depth look at the topic, see also F. MEDICO, *Il doppio custode*, in *Sociologia del diritto* (on-line), 2023, p. 1 ff., available at: <https://doi.org/10.30682/sg328>.

understood in the proper sense: AI does not merely support or speed up already existing activities, but creates the conditions for the structural transformation of decision-making, legal and administrative processes. It enables – i.e. makes possible – new forms of “governance”, categorisation and intervention, often removing them from the explicit control of the human being. This shift, only apparently neutral, raises numerous axiological and constitutional questions.

If, on the one hand, AI is celebrated for its ability to process large volumes of data in a very short time, offering seemingly optimised solutions to complex problems, on the other, it introduces a new rationality into public decision-making processes, based on performance and predictivity criteria, which risks circumventing or compressing the principles of equality, proportionality and inclusion.

In the absence of conscious regulatory structuring, the promise of efficiency may turn into a silent multiplier of inequalities, capable of consolidating systemic forms of exclusion under the veil of technical objectivity.

In the framework of the most advanced studies on the interaction between law and technology, AI emerges not as a mere tool, but as a true “*epistemic technology*”, capable of reshaping the very construction of social reality³. Mireille Hildebrandt has clarified how such systems do not merely manage data, but define in advance what is relevant, which subjects are recognised for legal purposes, and which attributes count in the assessment of reality⁴. In this perspective, the algorithm takes on the function of an implicit norm: it already operates upstream of decisions, defining the criteria for selecting, categorising and weighing data, and selects what can enter the space of the considerable and the judiciable.

In other words, the algorithm’s decision-making power does not only manifest itself in the executive phase of the judgement, but acts upstream in the prefiguration of the conditions that make the decision itself possible. Each decision-making model, in fact, does not merely provide computational outcomes, but organises the field of the relevant, defining semantic priorities, activating selective filters and delimiting what can be considered visible and relevant for the deliberative outcome⁵. In this function, it does not assume a prescriptive guise in the classical sense of the norm, but embodies a structural normativity: silent, opaque, but capable of affecting the trajectories of recognition and access to rights.

It has been observed how this normativity presents itself under the appearance of technical neutrality, concealing, in reality, value choices embedded in the design of the algorithm⁶. The selection of data, the construction of categories, and the definition of operational thresholds are operations that reflect cultural and political decisions, and which, if not adequately scrutinised, tend to stabilise existing inequalities, removing them from critical scrutiny and the possibility of opposition. For this reason, the algorithm cannot be considered a neutral entity: it operates as a discriminating device, which requires legal instruments capable of revealing its logic and guaranteeing its democratic controllability.

The operational dimension of algorithms directly affects the distribution of opportunities and forms of social recognition: they include or exclude, they attribute visibility or determine marginalisation, shaping the cognitive and value space within which meanings

³ M. HILDEBRANDT, *Smart Technologies and the End(s) of Law*, Cheltenham, 2015.

⁴ M. HILDEBRANDT, *Law as Information in the Era of Data-Driven Agency*, in *Modern Law Review*, 2016.

⁵ K. MARTIN, *Ethical Implications and Accountability of Algorithms*, in *Journal of Business Ethics*, 2018.

⁶ On this point see: S. VENKATASUBRAMANIAN, *Structural disconnects between algorithmic decision-making and the law*, in *International Review of the Red Cross Blog*, 2019; M. KRANZBERG, *Technology and History: 'Kranzberg's Laws'*, in *Technology and Culture*, 1986, p. 544 ff.

are attributed and subjective positions are constructed. In this sense, each algorithmic system is configured as an interpretative grid of an axiological nature, capable of orienting institutional and intersubjective dynamics well before a normative elaboration or formalised decision is made.

This shift of the cognitive-decisional axis, operated by algorithmic architectures, highlights how the power exercised by AI is not exhausted in punctual or circumscribed outcomes, but extends to the systemic configuration of the spaces of the possible. It is precisely in the capacity to orientate the parameters of evaluation in advance – and thus to delimit what can become the object of consideration, attribution or exclusion – that its ordering power manifests itself. In this context, the theme of algorithmic normativity can no longer be confined to the sphere of justice or public administration alone, but transversally invests all the sectors in which automated decisions contribute to structuring the conditions of access to rights, resources and social recognition. Hence the need to question the rationality underlying these systems not only in functional terms, but in their axiological, normative and performative dimensions.

In this sense, the normativity that emanates from algorithmic systems does not take the prescriptive form typical of classical sources of law – such as laws or regulations – but manifests itself as a structural and diffuse normativity, analogous to that which Michel Foucault recognised in knowledge-power devices: it does not directly impose behaviour, but conditions the very premises of the thinkable and the practicable⁷. Although clothed with the authority of mathematical objectivity, such normativity is actually the outcome of design, technical and cultural choices embedded in the structure of the code, which are often neither visible nor contestable by users or stakeholders. Precisely for this reason, the algorithm tends to naturalise pre-existing inequalities, inscribing them in computational processes that guarantee their systematic and silent reproduction⁸.

This awareness forces law to abandon a merely reactive or regulatory posture. The algorithm should not be seen as a neutral object to be “disciplined” but as a “*form of informational power*”, i.e. a mechanism that redistributes knowledge and opportunities, with direct implications on legal subjectivity and individual autonomy⁹. Law must therefore become a critical instrument of unveiling and direction, capable of questioning the epistemic assumptions of technology and opposing the apparent neutrality of efficiency with the criterion of public justifiability.

At European level, Regulation (EU) 2024/1689, known as the AI Act¹⁰, marks a fundamental step in the construction of a unified regulatory framework on artificial intelligence, inspired by the principles of accountability, transparency and protection of fundamental rights. The regulation introduces a risk-based approach, distinguishing between low, medium and high-risk systems, and imposing specific obligations for the latter such as rights impact assessment, transparent technical documentation and traceability of the entire life cycle of the system. However, the binary classification between “*high risk*” and “*not high*”

⁷ M. FOUCAULT, *Sorvegliare e punire*, Turin, 1976.

⁸ J. MONTEIRO, M. MARRAFON, *Legitimidade democrática na governança algorítmica*, in *Revista Direitos Fundamentais & Democracia*, 2024.

⁹ J. E. COHEN, *Between Truth and Power: The Legal Constructions of Informational Capitalism*, 2019.

¹⁰ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828

risk” may prove insufficient to capture the pervasive and transversal nature of algorithmic inequalities, which often manifest themselves even in contexts not formally labelled as critical.

A comparative look allows one to grasp the parallel evolution of AI law in other legal contexts. In the United States, the regulatory debate is focusing on the introduction of genuine “*algorithmic accountability*”, which imposes transparency and verifiability obligations on all automated systems employed in sensitive sectors¹¹. In Canada, the *Artificial Intelligence and Data Act* (AIDA)¹², originally part of Bill C-27 (2022), was supposed to establish a risk classification system and governance requirements for the development and responsible use of AI. However, the draft legislation lapsed in January 2025 with the early end of the federal legislature and the resignation of Prime Minister Trudeau¹³. Brazil also saw significant progress with *Projeto de Lei no. 2338/2023*, approved by the Senate on 10 December 2024 and currently under consideration by the Chamber of Deputies¹⁴.

These developments, although heterogeneous, converge on a shared vision of the need to make artificial intelligence compatible with the principles of the rule of law: transparency, accountability, inclusion and effective protection of fundamental rights. Within this framework, the comparison not only highlights normative convergences between legal systems, but also offers critical criteria for enhancing the capacity of law to act on an increasingly hybrid terrain, where the algorithm tends to silently replace formal normativity. It is precisely from this comparative analysis that a common thread emerges: law cannot confine itself to intervening once algorithmic decisions have already produced their effects,

¹¹ In 2023, the bill known as the *Algorithmic Accountability Act* (S.2892), sponsored by Senators Ron Wyden and Cory Booker, was introduced in the US Senate. The text included significant obligations for companies that develop or use automated decision-making systems, requiring algorithmic impact assessments, audits on the risk of *bias*, and the obligation to make transparent the criteria used for user profiling in sectors considered sensitive (such as healthcare, criminal justice, education and employment). The bill also envisaged the creation of a public register of the systems used, under the supervision of the Federal Trade Commission, and the possibility for citizens to access the automated decision-making logic that affects them.

However, the proposal did not make it through the committees and was not passed by the end of the 118th Congress, thus lapsing. Nevertheless, it sets a relevant precedent for the legal debate on *algorithmic accountability*, as it represents the first organic attempt in the United States to regulate the transparency and accountability of artificial intelligence systems in a structured manner. The text has also helped lay the foundations for future legislative interventions, both at federal and state level, and has fostered a broader reflection on the balance between technological innovation and the protection of fundamental rights in the digital society. *Algorithmic Accountability Act of 2023*, S.2892, 118th Congress. Available at: <https://www.congress.gov/bills/118th-congress/senate-bill/2892>.

¹² Government of Canada, *Artificial Intelligence and Data Act (AIDA) - Companion Document*, <https://ISED-ISCDCANADA.CA/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>.

¹³ Currently, the Canadian approach is articulated in local initiatives, such as Ontario's Bill 194, and non-binding instruments, including a voluntary Code of Conduct published by the federal government to promote ethical and transparent practices in the AI lifecycle.

¹⁴ Brazil now ranks among the most active countries in Latin America in the construction of a legal framework for AI, inspired by the European AI Act but enriched by specific elements aimed at strengthening the protection of fundamental rights and the responsibility of providers. PL 2338/2023 is distinguished by the introduction of guaranteed digital rights - such as the right to know when interacting with an AI system, the right to an explanation of automated decisions, to challenge and to non-discrimination - and by the adoption of a risk-based classification of systems, which distinguishes between excessive risk (prohibited, as in the case of social scoring) and high risk, subject to transparency, impact assessment, data management and human oversight obligations. Senado Federal do Brasil, *Projeto de Lei no. 2338/2023*, <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>.

but must help shape the initial conditions within which those decisions take form. The asymmetries that AI tends to amplify indeed require legal involvement in the very configuration of such systems.

2. Bias and vulnerability: automated discrimination in critical areas

The contemporary legal debate on the use of algorithms in decision-making processes focuses in particular on two problematic nodes: the determination of legal responsibilities and the structural risk of *bias* inherent in machine learning models. In a decision-making ecosystem increasingly mediated by automated systems, it becomes essential to precisely identify legally clear criteria for the allocation of liability, especially in the presence of erroneous or discriminatory outcomes. The increasing adoption of artificial intelligence in strategic areas – such as the labour market, the provision of public services or the assessment of individual rights – has in fact raised serious questions in terms of accountability and transparency, exposing asymmetries between the decision-making power of algorithms and the ability of those involved to exercise effective control over such processes¹⁵. In the context of automated decision-making, the issue of accountability is one of the central challenges for the rule of law in the digital ecosystem. Traditionally, accountability in technology has been traced back to the individual developers or technical teams responsible for the design and implementation of systems¹⁶. However, this approach today proves inadequate with respect to the structural and organisational complexity of artificial intelligence systems, the effects of which are distributed along an articulated chain of actors, stages and application contexts. In this perspective, the AI act marks an important evolution, proposing a model of shared and distributed responsibility, which involves not only suppliers, but also users, importers, distributors and – for high-risk systems – even third parties in charge of conformity assessment. This configuration reflects a pluralistic conception of liability, based on transparency, traceability, monitoring and risk management obligations, which must be fulfilled throughout the system's life cycle, from design to post-marketing¹⁷.

This evolution in the legal configuration of liability makes it all the more urgent to reflect systematically on the other major critical issue related to the use of artificial intelligence: that of algorithmic *bias*. If, on the one hand, the challenge is to ensure that the systems are controllable and legally accountable, on the other hand, it is necessary to question the nature of the data and logics that feed these systems, since it is precisely in these that the most insidious distortions lurk, capable of compromising the impartiality of automated decisions and profoundly affecting the effective protection of fundamental rights¹⁸.

The concept of algorithmic *bias*, if from a technical-scientific point of view it can be considered a statistical distortion or an accidental dysfunction of learning models, requires –

¹⁵ M. GÜNTHER, A. KASIRZADEH, *Algorithmic and human decision making: for a double standard of transparency*, in *AI & Society*, 2021, p. 375 ff.

¹⁶ G. COMANDÉ, *Intelligenza artificiale e responsabilità tra «liability» e «accountability». Il carattere trasformativo dell'IA e il problema della responsabilità*, in *Analisi Giuridica dell'Economia*, 2019, 18(1), p. 169 ff.

¹⁷ S. KLEANTHOUS, M. KASINIDOU, P. BARLAS, J. OTTERBACHER, *Perception of fairness in algorithmic decisions: future developers' perspective*, in *Patterns*, 2022.

¹⁸ A.E. WALDMAN, *Algorithmic Legitimacy*, in W. BARFIELD (ed.), *The Cambridge Handbook of the Law of Algorithms*, Cambridge Law Handbooks, Cambridge University Press, Cambridge, 2020.

from a legal and philosophical-normative perspective – a much deeper and structural analysis. Reducing *bias* to a technical anomaly is tantamount to defusing its political and legal significance: it is not the unforeseen effect of an error, but the consistent outcome of systems that learn from data steeped in historical stratifications, social hierarchies and systemic inequalities. *Bias*, in this sense, does not emerge when the algorithm fails, but precisely when it functions as designed: it takes the form of automated discrimination, which translates and reproduces pre-existing social asymmetries into computational form. Training datasets are never neutral: they reflect worldviews, cultural representations and power relations¹⁹.

From the point of view of fundamental rights, the algorithmic *bias* represents an indirect but concrete violation of the principle of non-discrimination, which is an inalienable pillar in any democratic system.

This principle is recognised and guaranteed both at the domestic constitutional level and at the supranational and international level. In modern constitutional democracies, it takes the form not only of a prohibition of arbitrary differential treatment, but also of a positive obligation for states to prevent and counteract all forms of systemic inequality, even when it manifests itself through technologically mediated mechanisms.

At European Union level, the principle is formally enshrined in Article 21 of the Charter of Fundamental Rights²⁰, which prohibits any form of discrimination based in particular on sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or other opinion, membership of a national minority, property, birth, disability, age or sexual orientation. This precept, of direct applicability, imposes on legislators and administrations a reinforced obligation to scrutinise the indirect effects produced by the adoption of digital technologies, even when such effects cannot be traced back to a discriminatory intent. Article 21 of the Charter, interpreted in light of the case law of the Court of Justice, in fact requires a substantive understanding of the principle of equality: the assessment must not stop at the formal neutrality of measures, but must extend to the actual consequences they produce within society²¹. This approach, which emphasises the material dimension of equality, implies that even apparently neutral technological tools may be subject to scrutiny where, in their concrete application, they generate a differential impact on vulnerable groups.

Scholarly work has highlighted how the interaction between algorithmic systems and European anti-discrimination law reveals a structural risk: harmful effects often emerge not as explicit violations of the prohibition of discrimination, but as collateral outcomes of complex decision-making architectures, which are difficult to detect without an analysis directed at the results they produce. For this reason, the prohibition set out in Article 21 requires, according to a significant strand of contemporary studies, that public authorities adopt assessment tools capable of capturing forms of inequality that do not appear in the formulation of the rule, but in its social operation²². In this sense, the obligation to scrutinise

¹⁹ As the US mathematician and data scientist C. O'Neil, "*algorithms are opinions embedded in code*". C. O'NEIL, *Weapons of Math Destruction*, New York, 2016.

²⁰ European Union, Charter of Fundamental Rights of the European Union, in OJEU C 326, 26 October 2012.

²¹ On this point see: Court of Justice of the European Union, judgment of 16 July 2015, Case C-83/14, *CHEZ Razpredelenie Bulgaria AD*, ECLI:EU:C:2015:480; Court of Justice of the European Union, judgment of 11 April 2019, Case C-397/18, *DW v Nobel Plastiques Ibérica SA*, ECLI:EU:C:2019:309.

²² Several authors have argued that a structural incompatibility exists between EU anti-discrimination law and the statistical fairness metrics developed in machine learning. While EU law requires a contextual and

does not amount to a merely procedural control, but to a form of substantive oversight requiring institutions to examine the mechanisms through which technologies may structure or amplify pre-existing inequalities.

This evolving interpretation of Article 21, which links the principle of non-discrimination to a positive duty to prevent indirect harmful effects, provides the logical foundation for the subsequent analysis of the European Convention framework. It allows one to understand how anti-discrimination protection within the European multilevel legal order operates through the control of the concrete outcomes generated by practices, procedures and ostensibly neutral criteria, regardless of whether they originate from public authorities or private actors. This focus on effects, rather than on subjective intent, finds a coherent counterpart in Article 14 of the European Convention on Human Rights (ECHR) and in the case law of the European Court of Human Rights on the relevance of indirect discrimination.

From this perspective, Article 14 of the ECHR prohibits any form of discrimination «in the enjoyment of the rights and freedoms set forth» in the Convention and, although ancillary in nature, requiring a connection with another Convention right²³, it has progressively acquired substantive weight in the Court's jurisprudence. The Court has clarified that the prohibition does not concern only intentional disparities, but also requires the examination of the effects produced by rules, practices or ostensibly neutral mechanisms where they systematically or foreseeably affect specific social groups. Within this interpretative framework, the Court has recognised the full legal relevance of indirect discrimination, qualifying as potentially incompatible with Article 14 all situations in which a disadvantage resulting from the uniform application of a measure translates, in practice, into a differential impact lacking an objective and reasonable justification²⁴.

Numerous empirical studies confirm the urgency of addressing the legal implications of algorithmic *bias* in areas with a high impact on fundamental rights. In healthcare, extensive research by Ziad Obermeyer et al. published in 2019 in *The New England Journal of Medicine* found that an algorithm widely employed by a major US insurance company to identify

substantive assessment of the differential effects produced by a given measure, technical metrics operate through standardised and abstract criteria that are unable to capture the complexity of indirect discrimination as recognised by the Court of Justice. In algorithmic systems, the social cues typically associated with human discrimination (such as stereotypes, prejudices or observable behaviours) are absent, which makes the legal identification of disparate treatment considerably more complex. On this basis, it has been proposed to adopt a baseline statistical measure, *conditional demographic disparity* (CDD), which may serve as a minimal reference point for detecting automated disparities while preserving the interpretative margin required by the European model of non-discrimination. S. WACHTER, B. D. MITTELSTADT, C. RUSSELL, *Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI*, in *arXiv* (online), 2020, arXiv:2005.05906.

²³ European Court of Human Rights, *Guida all'articolo 14 della Convenzione*, updated edition 31 December 2020. Available:

www.giustizia.it/cmsresources/cms/documents/guida_cedu_articolo14_e_articolo1_protocollo12.pdf.

²⁴ A paradigmatic case is represented by the Grand Chamber judgment *D.H. and others v. Czech Republic*, in which the Court condemned the practice of systematically placing Roma children in special schools on the basis of psycho-aptitude tests that were formally neutral, but which were substantially discriminatory against that community (ECHR, Grand Chamber, *D.H. and others v. Czech Republic*, Appl. No. 57325/00, judgment of 13 November 2007, ECLI:CE:ECHR:2007:1113JUD005732500). In *Opuz v. Turkey*, the Court stated that the inaction of public authorities - in that case the absence of adequate protective measures towards a woman victim of domestic violence - may also constitute discrimination under Art. 14 ECHR, if it produces asymmetrical consequences to the detriment of vulnerable groups, such as women, even in the absence of a direct discriminatory intention (ECHR, *Opuz v. Turkey*, Rec. no. 33401/02, judgment of 9 June 2009, ECLI:CE:ECHR:2009:0609JUD003340102).

patients in need of complex care systematically assigned lower risk scores to African-American patients than to white patients with identical medical conditions²⁵. The system, in fact, used past health expenditure as a proxy variable, believing it to be indicative of clinical need. However, this parameter proved to be severely distorting, as it reflected systemic inequalities in access to care by African American communities, historically excluded from adequate and continuous health care pathways. This distortion was not an accidental technical error, but the predictable outcome of a decision-making architecture that, by selecting an economic metric as the criterion of relevance, reproduced – rather than corrected – the racial asymmetries already present in the healthcare system.

Other recent studies have highlighted how deep learning models applied to radiological diagnostics tend to perform significantly poorer for ethnic minority patients²⁶. In particular, several automatic classification algorithms used on radiographic images were analysed and found that their accuracy was systematically higher for white patients than for black, Hispanic or Asian patients. This discrepancy is not accidental: it depends on unbalanced training datasets, in which images of white patients are over-represented, and on metrics optimised on majority populations. This leads to a systemic underdiagnosis affecting specific groups, thereby compromising equitable access to care and potentially infringing on the right to health²⁷.

The use of proxies that are not validated from a social equity perspective shows how the selection of algorithmic indicators is not a neutral operation, but an axiological and normative choice that directly affects the structure of rights²⁸.

In the context of the criminal justice process, the introduction of artificial intelligence systems to support judicial decisions – e.g. in assessing the risk of recidivism, the allocation of pre-trial detention or the determination of alternative measures – has raised growing questions about the respect of constitutional and conventional guarantees, such as the principle of due process, the right of defence and the prohibition of discrimination. A case in point is the use of predictive algorithms such as COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), used by several US courts to estimate the probability of a defendant committing a new crime. Independent studies have revealed a systematic tendency for the algorithm to classify as ‘high risk’ African-American subjects who, in fact, would not have committed crimes after the assessment. In contrast, white defendants with a similar profile were more frequently classified as ‘low risk’ even in the presence of significant prior history or comparable past behaviour²⁹. This predictive *bias* was not due to technical error, but reflected a decision-making structure based on historical data contaminated by systemic biases – such as selective policing policies and disparities in access to alternative measures –

²⁵ Z. OBERMEYER, B. POWELL, C. AUFMAN, S. VASAN, *Dissecting racial bias in an algorithm used to manage the health of populations*, in *New England Journal of Medicine*, 2019, vol. 381, p. 2471 ff.

²⁶ L. SEYYED-KALANTARI, H. ZHANG, M.B.A. MCDERMOTT, I.Y. CHEN, M. GHASSEMI, *Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations*, in *Nature Medicine*, 27, 2021, p. 2176 ff.

²⁷ *Ibid.*

²⁸ Health data are never neutral or purely numerical elements: they always implicitly reflect the social structures and value hierarchies that define the meaning and treatment of human lives. Every piece of clinical data collected, every indicator used in healthcare, in fact incorporates political and cultural choices that influence what is measured, for whom it is measured and for what purpose. In this sense, the very construction of the data is already a normative act, reflecting institutional priorities and prior power relations. V. EUBANKS, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, New York, 2018.

²⁹ J. ANGO WINNER, J. LARSON, S. KIRCHNER, *The Machine Bias*, in *ProPublica* (online journal), 2016.

which the algorithm replicated, codified and reiterated in its output³⁰. Such biased predictions, although formally neutral, resulted in asymmetrical judicial treatment, influencing the granting of probation, bail or alternative sentences. The proprietary nature of the software, protected by industrial secrecy, has long prevented defendants from accessing the algorithm's operating criteria, making effective judicial review of the underlying decision-making mechanisms impossible³¹.

The problem, however, is not limited to the quality of the data³² or statistical errors. In a broader perspective, it is necessary to consider the active role of artificial intelligence in the construction of social reality: algorithmic technologies do not merely describe the world, but help to shape it on the basis of narrative, cultural and power structures embedded in data and models³³. The algorithm operates as a performative device, capable of selecting what is to be visible, categorised or measured, thus contributing to the definition of standards of relevance, deservingness and legal recognition. This constitutive character has been highlighted by authoritative studies that have highlighted how algorithmic systems, far from being neutral instruments, interact with social structures and contribute to their reproduction³⁴. The technical infrastructure that supports AI incorporates worldviews, values and political aims that are often neither transparent nor accessible to those involved in automated decision-making processes. In doing so, the algorithm can reinforce pre-existing social hierarchies, making them part of a codified symbolic order that is difficult to challenge.

This performative capacity of the algorithm highlights a further criticality: the emergence of forms of automatic classification that tend to crystallise individuals into opaque and unchangeable groups, with no possibility of reconfiguration or exit. This is the phenomenon described by Sandra Wachter as *algorithmic immutability*, according to which the sets constructed by AI systems – although not based on intrinsic characteristics such as race or gender – end up behaving, in decision-making practice, as if they were fixed and unchallengeable attributes³⁵.

This algorithmic crystallisation produces an effect similar to that of discrimination based on protected categories: it impedes social mobility, reduces the possibility of rectification or explanation and restricts individual autonomy, effectively excluding individuals from fundamental opportunities such as work, credit, education or health care.

³⁰ C. THOMAS, A. NUNEZ, *Automating Judicial Discretion: How Algorithmic Risk Assessments in Pretrial Adjudications Violate Equal Protection Rights on the Basis of Race*, in *Law & Inequality*, 2022, p. 371 ff.

³¹ *Ibid.*

³² Data quality is a central aspect in the construction of algorithmic systems, as it directly influences the reliability and fairness of automated decisions. A quality dataset is not simply an ordered set of information, but a product constructed through methodological, epistemological and often also political choices. The main requirements that guarantee its suitability include completeness (i.e. the ability to adequately represent all relevant subjects with respect to the application domain), accuracy (the correspondence of the data to empirical reality), timeliness (the use of up-to-date information), representativeness (the balanced presence of the different social categories), and the absence of systemic biases (preventing stereotypes or historical prejudices from being encoded in the data). P. CHEN, L. WU, L. WANG, *AI Fairness in Data Management and Analytics: A Review on Challenges, Methodologies and Applications*, in *Applied Sciences*, 2023, 13(18), p. 10258 ff.

³³ K. CRAWFORD, *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven, 2021.

³⁴ N. SEAVER, *Algorithms as culture: Some tactics for the ethnography of algorithmic systems*, in *Big Data & Society* (online journal), 2017, p. 1 ss; C. O'NEIL, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York, 2016; J. NIDA-RÜMELIN, N. WEIDENFELD, *Digitaler Humanismus. Eine Ethik für das Zeitalter der Künstlichen Intelligenz*, München, 2018.

³⁵ S. WACHTER, *The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law*, in *SSRN Electronic Journal* (online journal), 2022.

Classification is thus transformed into a technical predestination, which operates silently but systemically, consolidating new forms of inequality not recognised by the current anti-discrimination legal architecture.

3. *Towards an inclusive algorithmic normativity: law between substantive justice and computational design*

The growing affirmation of algorithmic technologies in decision-making processes poses numerous questions for legal theory: is law still able to exercise an effective regulatory function in an ecosystem in which decisions are increasingly entrusted to opaque, dynamic and self-learning models?

Computational architectures, in fact, do not merely perform support functions, but are configured as regulatory devices in their own right, capable of affecting social relations, redistributing power and determining access, exclusions and recognitions. In this perspective, law is called upon to overcome a reactive and formalistic conception of legality in order to assume a design posture, capable of intervening *ex ante* in the definition of algorithmic logics, according to a paradigm of legal protection integrated in the design phase³⁶.

This approach implies an epistemological reversal: law can no longer be read as a separate system with respect to technology, but must become an integral part of its infrastructure. The aim is to prevent the executive force of the code from taking the place of the normative force of the law, making automated decisions transparent, contestable and respectful of fundamental rights³⁷.

A legally pregnant analysis of algorithmic performativity requires abandoning a technicist and static vision of the code, to embrace a relational and dynamic paradigm, in which algorithms are understood as “*socio-technical assemblages*” whose normativity stratifies over time. It is in this perspective that the concept of the “*biography of algorithms*”, elaborated within the framework of organisation theory to describe the evolutionary trajectory of algorithms as socio-technical artefacts, fits in. According to this approach, every algorithmic system is structured through a progressive stratification of “*programmes of action*” – worldviews, operational metaphors and performative expectations – that guide its functions and condition its effects, often on the basis of ideological assumptions or implicit structural interests³⁸. This, from a critical perspective, raises essential questions about rule of law in digital environments. If, in fact, the algorithm is the bearer of embedded logics and values, and if these logics influence decisions that affect fundamental rights, then the very architecture of the code must be subject to legal scrutiny. The apparent neutrality of the machine – the result of an iterative process of ideological and technical “*layering*” – risks concealing the concentration of power in the hands of technical or commercial actors who

³⁶ M. HILDEBRANDT, *Smart Technologies and the End(s) of Law*, cit., p. 34.

³⁷ M. HILDEBRANDT, *Legal Protection by Design: Objections and Refutations*, in *Legisprudence*, vol. 5, no. 2, 2011.

³⁸ During its lifecycle, an algorithm does not merely perform predetermined functions, but incorporates implicit normative representations: models of the intended user, visions of organisational efficiency, selection logics of relevant information. These - often invisible - inscriptions guide the translation of abstract goals (e.g. security, productivity, merit) into technical computational mechanisms that, in turn, rewrite the criteria of access, evaluation and recognition. It is in this sense that algorithmic design is configured as an eminently legal moment, in which the spaces of individual freedom, procedural transparency and substantive equality are 'programmed'. S. GLASER, A. BIJL, P. STORM, *The Biographies of Algorithms: Understanding the Performative Dynamics of Algorithmic Systems in Organisations*, in *Organisation Theory*, vol. 2, no. 3, 2021, pp. 1-21.

are not democratically accountable. Hence the urgency of a legal gaze that is both epistemological and structural, capable of interrogating the design process as the *locus* of contemporary normativity.

The effective inclusion of fundamental legal principles in computational mechanisms is an unavoidable condition to ensure that artificial intelligence develops and operates in compliance with the canons of constitutional legality and fundamental human rights. In a historical phase in which normativity tends to migrate from legislative text to source code, law can no longer limit itself to an *ex post* control function, but must claim a constitutive role in the definition of digital architectures. Algorithms do not merely support decision-making; they prefigure, structure and, in many cases, determine decisions in a prescriptive manner. This occurs in strategic domains such as personnel selection, creditworthiness assessment or the distribution of information on digital platforms, where code becomes a genuine normative infrastructure³⁹. In automated CV-screening processes, for example, the algorithm determines which candidates progress to the next stage: the system does not simply assist HR, but effectively prescribes the ordering of opportunities. Empirical studies confirm that candidates filtered by AI systems are subsequently hired at significantly higher rates than those selected through traditional methods, demonstrating the performative power of algorithmic mechanisms over occupational mobility⁴⁰. Similarly, in credit scoring models, the automated attribution of a score determines access to, or exclusion from, essential services such as mortgages, loans or credit lines. Such decisions appear neutral and data-driven, yet they embed and amplify pre-existing social vulnerabilities, particularly for individuals without established financial histories or from migrant backgrounds⁴¹. On digital platforms, moreover, algorithmic ranking exercises an even more radical normative function: it determines which content becomes visible and which remains invisible, thereby shaping the distribution of economic opportunities (monetisation), access to information and even the formation of public opinion⁴². As extensively documented in digital sociology and information science, these systems do not simply reflect user preferences, but prescribe what becomes salient, redefining hierarchies of value and visibility⁴³.

These examples reveal a structural truth: algorithmic normativity operates upstream, in the phase of study and design, and it is precisely at this stage that law must intervene in a constitutive manner, no longer as a mechanism of post-hoc correction. If code determines who sees what, who gains access to which opportunities, and who is deemed “*reliable*” or “*deserving*”, then the algorithmic architecture becomes the new terrain on which the conditions of possibility for social action are articulated. To remain faithful to its guarantee function, law must therefore act within the algorithmic environment, shaping its evaluative premises and ensuring that conditions of access and recognition remain consistent with the principles of the rule of law.

³⁹ L. LESSIG, *Code: And Other Laws of Cyberspace*, Basic Books, New York, 1999.

⁴⁰ M. RAGHAVAN, S. BAROCAS, J. KLEINBERG, K. LEVY, *Mitigating bias in algorithmic hiring: evaluating claims and practices*, in «Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT '20)», Association for Computing Machinery, New York, 2020, pp. 469-481.

⁴¹ M. HURLEY, J. ADEBAYO, *Credit Scoring in the Era of Big Data*, in *Yale Journal of Law and Technology*, vol. 18, 2017, pp. 5 ss.

⁴² T. GILLESPIE, *Custodians of the Internet. Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*, New Haven–London, Yale University Press, 2018.

⁴³ On this point see: R.K. NIELSEN, S.A. GANTER, *The Power of Platforms. Shaping Media and Society*, New York, Oxford University Press, 2022; T. GILLESPIE, *Custodians of the Internet. Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*, New Haven–London, Yale University Press, 2018, ISBN 9780300173130.

In this perspective, the shift from a merely prescriptive conception of law to a co-design paradigm is not a rhetorical formula, but a structural requirement clearly articulated in the Ethics Guidelines for Trustworthy AI of the High-Level Expert Group on AI. The Guidelines identify three inseparable dimensions of trustworthy artificial intelligence: lawfulness, ethical soundness and technical robustness⁴⁴. These requirements do not operate as independent layers, but as components that must be integrated into system design from the earliest stages. The configuration environment of decision-making structures can therefore no longer be conceived as an exclusively technical space, but as one in which legal knowledge contributes to defining limits, orienting criteria and structural conditions of operation.

Within this framework, the contribution of the jurist does not consist in imposing an already-formed regulatory framework from the outside, but in translating fundamental legal categories, such as proportionality, non-discrimination, dignity and transparency, into constraints and parameters that shape design choices. This occurs, for example, in the selection of relevant variables, the definition of optimisation metrics, the calibration of algorithmic sensitivity thresholds or the establishment of effective mechanisms for contestability. In this sense, legal intervention takes the form of a conceptual mediation that brings into the computational environment the principles that sustain constitutional legality.

System design must be oriented towards ensuring contestability, transparency and meaningful avenues for review. In the European Guidelines, these elements are not ethical accessories but essential conditions for the legitimacy of any automated system. This entails building technologies that allow users to understand decision-making logics, verify the proportionality of interventions and influence the ways in which their data are processed and their positions determined. In this way, legal protection does not operate solely a posteriori, correcting errors, but acts upon the very conditions of decision-making, shaping the technological environment so that it reflects the fundamental principles of the rule of law.

A particularly illuminating reflection in this regard comes from the European research project COHUBICOL – *Counting as a Human Being in the Era of Computational Law*, directed by Mireille Hildebrandt and funded by the European Research Council. The investigation conducted within COHUBICOL has shown how the transposition of law into computational form – in both *data-driven* and *code-driven* models – profoundly transforms the very nature of legal normativity. Law, from being a system of rules open to interpretation, dissent and contestation (and thus conforming to the Rule of Law), runs the risk of becoming a closed, opaque and self-executing system, where the rule is inscribed in automatic procedures that admit neither exceptions nor interpretative pluralism⁴⁵.

⁴⁴ The High-Level Expert Group on AI of the European Commission articulated the notion of “*trustworthy AI*” around three core requirements: lawfulness, understood as conformity of the system with the legal order; ethics, conceived as respect for fundamental European values such as dignity, autonomy, justice and non-discrimination; and technical and societal robustness, which includes resilience, safety, risk management and human oversight. The Guidelines further emphasise the *by design* character of these requirements: they are not to be verified *ex post*, but must be embedded in the design choices concerning data, models, metrics and mechanisms of contestability. Particular importance is also attributed to accountability, traceability and explainability of outputs, regarded as necessary instruments for ensuring that legal principles operate as structural constraints rather than as mere ethical aspirations.

HIGH-LEVEL EXPERT GROUP ON AI, *Ethics Guidelines for Trustworthy AI*, European Commission, 2019.

⁴⁵ The concept of *law as computation*, which is at the heart of the project, takes two main forms: on the one hand, *data-driven law*, based on the use of machine learning techniques to predict or suggest the application of rules;

In this context, the principle of contestability – i.e. the right of every individual to understand, discuss and challenge decisions that affect him or her – appears threatened by the opaque automation of decision-making processes. As pointed out by part of the doctrine, the computational codification of law tends to flatten the deliberative and dialectical dimension of law itself, replacing *practical reason* with *technical rationality* and reducing legal subjectivity to an algorithmic function⁴⁶. The effect is structurally political: the legitimacy of the decision no longer lies in its public justifiability, but in its procedural efficiency.

In a decision-making ecosystem increasingly permeated by computational opacity, formal equality proves to be a legally inadequate instrument. Article 3(2) of the Italian Constitution in fact imposes a positive obligation on the state: it is not enough to avoid discriminatory treatment, but action must be taken to actively remove economic and social obstacles that effectively limit the freedom and participation of citizens. Applied to the context of algorithmic technologies, this principle implies the need to critically investigate the operating logics of automated systems, making them transparent, accessible and subject to democratic control.

Algorithmic rationality, precisely because it acts *ex ante* by selecting, categorising and attributing relevance, must be evaluated not only in terms of technical efficiency, but especially with respect to its ability to promote or hinder inclusion⁴⁷. A constitution that recognises equality as a dynamic and substantive principle also requires technological artefacts to be judged according to their impact on distributive justice and effective access to rights.

European Union law already offers a first regulatory basis for containing the risks associated with automated decision-making. Article 22 of Regulation (EU) 2016/679 (*General Data Protection Regulation – GDPR*)⁴⁸ grants the data subject the right not to be subjected to decisions based solely on automated processing, when such decisions produce legal effects or significantly affect the individual. If interpreted substantively, this provision does not merely prohibit automated decision-making, but also imposes an obligation to ensure

on the other hand, *code-driven law*, which takes the form of the automation of legal rules in the form of executable code – think, for instance, of *smart contracts* on *blockchain* platforms. Both approaches, if not properly governed, risk turning law into a closed, deterministic and opaque device, unable to guarantee transparency, contestability and interpretability.

Hildebrandt points out the dangers associated with such a transformation, emphasising how the computation of law can compress the individual's ability to challenge decisions that affect him, reduce the interpretative flexibility that characterises legal argumentation and, above all, shift the centre of gravity of normative power from public and democratically legitimised institutions towards private actors, such as technology companies and code programmers. Faced with this scenario, the COHUBICOL project proposes the development of a *new computational legal hermeneutics*, capable of integrating tools of classical legal theory with technical computer skills. The aim is to preserve the fundamental principles of legality – publicity, predictability, accountability – even within automated digital ecosystems. In this sense, law must not passively adapt to technology, but must know how to 'write' it according to its own principles: this is the profound sense of the notion of *legal protection by design*. EUROPEAN COMMISSION, *Cordis - Cobubicol project fact sheet counting as a human being in the era of computational law*, available at: <https://cordis.europa.eu/project/id/788734>.

⁴⁶ P. HACKER, *Explainable AI under Contract and Tort Law: Legal Incentives and Technical Challenges*, in *Artificial Intelligence and Law*, 29, 2021.

⁴⁷ L. VIOLINI, *Costituzione, tecnologie e inclusione: verso una cittadinanza digitale sostanziale*, in *Costituzione e Inclusione Digitale*, Giappichelli, 2022.

⁴⁸ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation), published in OJEU. L 119 of 4 May 2016

genuine, effective and informed human control, which is not merely a formal ratification of the algorithmic output⁴⁹.

However, the scope of Article 22 is both circumscribed and insufficient. The rule protects the individual only in the presence of direct and relevant consequences, potentially leaving a vast grey area uncovered: that in which algorithms do not produce immediate legal effects, but nevertheless contribute, right from the design and profiling stage, to generating structural asymmetries and systemic discrimination. In this perspective, the provision of the GDPR proves incapable of addressing the forms of exclusion that are rooted in the very design of automated decision-making systems, making it necessary to integrate it with substantive principles of algorithmic justice and inclusion.

It is precisely in this direction that a deeper reflection on the role of law is required: the construction of a genuinely inclusive algorithmic normativity cannot stop at the mere enunciation of principles of non-discrimination nor be satisfied by an ethical supervision entrusted to soft law instruments or to the deontological responsibility of developers⁵⁰. Such an approach – although important on a cultural level – has proved insufficient to guarantee effective protection of fundamental rights in digital contexts, where inequalities no longer manifest themselves only in the blatant violation of norms, but creep into the technical assumptions that structure automated decisions. The challenge facing contemporary legal culture is therefore more radical: it is necessary to go beyond the traditional horizon of law as a discipline of behaviour, in order to recognise its role as an enabling – or limiting – architecture of the conditions of possibility of social action.

It is in this perspective that regulatory intervention must address not only the explicit contents of algorithms, but their very design, training datasets, optimisation metrics, and verification and validation methods. The Italian case of the automated system for the management of the Citizenship Income⁵¹ – which generated discriminatory exclusions on

⁴⁹ This reading has also been transposed by national case law: consider, for example, ruling No. 8472/2019 of the Council of State, which deemed illegitimate the use of an algorithm for the automatic allocation of school seats, in the absence of a real margin of discretionary assessment by the administration⁴⁹. The administrative judge clarified that entrusting the decision entirely to the algorithm empties the procedure of its human dimension, violating the principle of good conduct and impartiality (art. 97 of the Constitution), as well as the participatory rights of the recipient of the measure. Sent. Cons. Stato, Sec. VI, 13 December 2019, no. 8472.

⁵⁰ European governance of artificial intelligence was initially based on soft law instruments and non-binding ethical principles without direct legal effect. The Ethics Guidelines for Trustworthy AI published in 2019 by the High-Level Expert Group on AI promote values such as transparency, justice, accountability and non-discrimination, but do not provide for coercive obligations or enforcement mechanisms. In parallel, even voluntary international instruments – such as the Toronto Declaration (2018) or the Hiroshima Code of Conduct (2023, promoted by the G7) – while offering important value references, do not by themselves guarantee effective protection of fundamental rights, especially with respect to the systemic biases embedded in algorithmic systems.

⁵¹ The management system for the Citizenship Income, used in Italy as of 2019, was also based on automated mechanisms for assessing the 'adequacy' of applications, including algorithms for profiling beneficiaries and *ex ante* controls. Some critical studies have denounced the use of opaque criteria not accessible to citizens and not verifiable in court, which would have excluded individuals in a socially vulnerable condition without the possibility of cross-examination. The case raised important questions in terms of transparency, accountability and human control over automated decisions.

M. DE ANGELIS, *Vulnerabilità e misure di sostegno al reddito in Italia: il caso del reddito di cittadinanza*, in *Economia e società regionale*, no. 1/2023, pp. 116-136.

the basis of opaque and unchallengeable criteria – and the AMS algorithm⁵² in Austria, which introduced penalising forms of profiling for categories already vulnerable in the labour market, represent two paradigmatic examples of the systemic risk of exclusion that is generated not by discriminatory intentionality, but by the naturalisation of pre-existing structures of inequality. In such cases, the algorithm functions as a device for consolidating the existing social order, through a computational rationality that interprets reality through historically stratified statistical regularities, without questioning its ethical or constitutional implications.

Far from being attributable to a programming error, the discriminatory risk then derives from an epistemic and axiological deficit that invests the entire life cycle of the system: from the engineering of the model to the absence of public control. For this reason, inclusion cannot be considered a final attribute to be verified a posteriori, but must become a structural parameter of the algorithmic design, the object of strict legal regulation and democratic accountability.

The AI Act represents an important attempt at a response, introducing a binding framework based on risk assessment and shared responsibility throughout the systems' life cycle. The provisions on high-risk systems, in particular Articles 29 and 64, provide for tools such as fundamental rights impact assessments and independent audits, which aim to ensure a form of *prior* and participatory *accountability*. However, the binary high/high-risk categorisation is insufficient to intercept the structural and ubiquitous nature of computational inequalities.

Particular attention also deserves the role of sectoral codes of conduct provided for in Articles 29 and 95, which – although being soft law instruments – can be an effective lever for algorithmic inclusion, provided they are co-designed with the active involvement of impacted communities. In this sense, algorithmic fairness is measured not only in the statistical correctness of results (*fairness by metrics*), but in the democratic quality of decision-making processes and in the possibility for citizens to understand, challenge and influence the logics that govern their daily lives.

In this framework, the idea of justice cannot be reduced to the procedural fairness of the algorithm: it must also include the possibility for citizens to understand, contest and influence the decisions that affect them. As Aristotle recalled in the *Nicomachean Ethics*, equity (*epieikeia*) is what enables law to adapt to the variety of concrete situations, where the general rule proves insufficient⁵³. Translated into the context of artificial intelligence, this means constructing a *computational legality* that does not abdicate the technicalities of algorithmic rationality, but critically traverses them, orienting them towards constitutionally compatible

⁵² Austria's AMS (Arbeitsmarktservice) system, introduced in 2018, employed a predictive algorithm to classify the unemployed into three 'employability' categories, based on variables such as age, gender, nationality, level of education and work history. This classification resulted in differential access to public employment services, particularly penalising women, migrants and people with disabilities. The project drew strong criticism from civil society, the academic community and the Austrian Data Protection Authority, which considered the system to be non-compliant with the GDPR due to violation of the principle of fairness and lack of transparency. In 2020, its implementation was suspended. D. ALLHUTTER, F. CECIL, F. FISCHER, G. GRILL, A. MAGER, *Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective*, in *Frontiers in Big Data*, vol. 3, 2020.

⁵³ Aristotle defines equity as the corrective to universal law when the latter, due to its generality, is inadequate with respect to the complexity of the concrete case. Equity, as superior and flexible justice, assumes a central role in complex regulatory contexts, such as today's dominated by automated decision-making systems. ARISTOTELE, *Etica Nicomachea*, transl., intr. and notes by C. NATALI, Laterza, Roma-Bari, 1999.

ends. It is in this perspective that the jurist is called upon not only to regulate AI, but to co-design it, translating the principles of substantive equality into normative design criteria capable of preventing injustice before it manifests itself.

4. *Conclusions*

The growing centrality of artificial intelligence in public and private decision-making processes calls for a profound reconfiguration of the role of law. In a context in which normativity tends to migrate from the text to the computational structure, it is no longer sufficient to preside *ex post* over the effects of automated decisions: it is necessary to preside *ex ante* over the architectures that generate them. In other words, the jurist can no longer limit himself to controlling the legality of algorithmic decisions: he must participate in their semantic and axiological design.

This change requires a new legal paradigm: law no longer as a reactive system, but as an embedded normative infrastructure, capable of directly affecting the design of AI systems. This is achieved concretely through at least three strategic directions:

1. Constitutionalisation of algorithmic design: fundamental principles – equality, transparency, due process – must become computable requirements, i.e. technical specifications embedded in datasets, optimisation metrics and variable selection criteria. Inclusion can no longer be a criterion assessed downstream, but must become a structural design constraint.
2. Structural and cyclical accountability: there is a need for a legal model of distributed accountability, which accompanies the system throughout its life cycle, imposing audit, documentation and explainability obligations not only at the time of placing on the market, but also at the time of updating, reconfiguration or migration of systems. Law must act as an active memory of the algorithmic cycle, not as a footnote to technology.
3. Regulatory co-production between law and technology: the jurist must participate in development processes, adopting collaborative methodologies (co-design) with computer scientists, engineers, sociologists and impacted communities. In this perspective, tools such as *ex ante* fundamental rights impact assessments, in-house ethical-legal committees and user-accessible dispute platforms become new-generation legal devices, capable of anticipating conflict, not just mitigating its effects.

It is therefore not a question of “*humanising*” the algorithm with declarations of principle, but of legalising its operational logic. The law is called upon to exercise a new form of vigilance: no longer just on the legitimacy of the act, but on the grammar of the code. From this perspective, legality is no longer a constraint external to the technical system, but a design criterion. Algorithmic fairness, therefore, is not the fortuitous outcome of a good technical operation, but the result of a legal-political choice, embedded in the very form of the automatic decision.

Only in this way can law continue to exercise its historical function: not that of chasing the transformations of technology, but of orienting them towards the horizon of constitutional justice.

At its core, the task of law in the algorithmic age is not merely to police the boundaries of technology, but to preserve the space of subjectivity and democratic deliberation. Justice is not an emergent property of computation, but a political project that requires constant vigilance, even when decisions are filtered or reorganised by computational systems. Otherwise, there is a risk that logics of automatic efficiency will replace forms of public evaluation grounded in contestation, justification and responsibility. Law thus remains the forum in which these transformations must be interpreted, debated and rendered compatible with human dignity and with the substantive equality that democratic constitutions commit to guaranteeing.



PAOLO DALLE CARBONARE*

DOVERE TRIBUTARIO E INTERESSE FISCALE: LA DIFFICILE SOPRAVVIVENZA DEL DIRITTO ALLA RISERVATEZZA NELL'ERA DELL'INTELLIGENZA ARTIFICIALE

SOMMARIO: 1. Introduzione. – 2. Potere fiscale e diritto alla riservatezza del contribuente. – 3. La bulimica rincorsa alla raccolta di dati: un digiuno di diritti per il contribuente? – 4. Sorveglianza algoritmica: il colpo di grazia ad un diritto agonizzante. – 5. Conclusioni.

1. *Introduzione*

Il potere di imporre tributi è un elemento costitutivo della sovranità dello Stato attraverso il quale sono acquisite le risorse per la sopravvivenza della collettività. Tale sovranità si manifesta attraverso atti autoritativi con cui lo Stato impone ai privati il prelievo di risorse da impiegare per il soddisfacimento di interessi collettivi e la garanzia dei diritti costituzionali¹. Visto dalla prospettiva del cittadino, l'interesse dello Stato si traduce in un dovere inderogabile di concorrere al sostenimento della spesa pubblica quale manifestazione concreta del dovere di solidarietà previsto dall'articolo 2 della Costituzione.

Il pagamento dei tributi non è, pertanto, solo un obbligo giuridico, ma soprattutto la manifestazione concreta del patto sociale che sostiene lo Stato costituzionale. Interesse pubblico e dovere privato si intrecciano, trovando il loro punto di equilibrio nel principio di capacità contributiva, che rappresenta un limite sia qualitativo che quantitativo all'imposizione fiscale.

In quest'ottica, il dovere fiscale assume rilevanza non tanto in termini di soggezione, pur essendo regolato e sanzionato dalla legge, quanto piuttosto come strumento essenziale

* Dottorando di ricerca in Scienze giuridiche e politiche, Università degli studi "Guglielmo Marconi".

¹ Per un interessante approfondimento sul tema del potere tributario vedasi P. Boria, *Il potere tributario: politica e tributi nel corso dei secoli*, Bologna, 2021 e F. Amatucci, *La sovranità fiscale nazionale – Tra erosione e consolidamento*, Torino, 2022.

per garantire le risorse necessarie all'erogazione dei servizi pubblici, alla tutela degli altri diritti fondamentali² e alla redistribuzione del reddito tra i consociati³.

L'attività di gestione dei tributi svolta dallo Stato attraverso la sua amministrazione comporta talvolta la trasformazione dell'interesse fiscale in un valore preminente e assoluto, nel quale la ricerca del gettito erariale rischia di comprimere gli stessi diritti costituzionali che è chiamata a garantire, quali diritto alla riservatezza, alla difesa, libertà di iniziativa economica, la tutela della proprietà ed il buon andamento e imparzialità dell'amministrazione. Tali limitazioni si estrinsecano attraverso l'accesso eccessivo ed indiscriminato ai dati dei contribuenti, l'imposizione di presunzioni legali inique, inversione dell'onere della prova, limitazioni alla tutela giurisdizionale, procedure aggressive di riscossione. Quando il legislatore e l'amministrazione si spingono a comprimere le libertà del cittadino, si allontanano «dalle altissime ragioni di civiltà giuridica che fondano il dovere tributario, con gravi conseguenze in termini di disorientamento non solo dello stesso sviluppo dell'ordinamento, ma anche del relativo contesto sociale»⁴.

2. Potere fiscale e diritto alla riservatezza del contribuente

Il bilanciamento tra potere fiscale e diritti del contribuente appare particolarmente complesso con riferimento al diritto alla riservatezza del contribuente, oggi minacciato dal reclutamento, da parte del sistema fiscale, degli strumenti di intelligenza artificiale⁵.

Il diritto alla riservatezza della persona, inteso quale manifestazione del diritto fondamentale all'intangibilità della sfera privata⁶, pur non ritrovando un espresso riferimento nella Costituzione, può essere dedotto dalla lettura degli articoli 2, 14 e 15 (diritti inviolabili dell'uomo, tutela del domicilio e della segretezza della corrispondenza). Il diritto è stato più

² Sulla funzione di garanzia dei diritti fondamentali da parte delle tasse suggestiva è la visione offerta da S. HOLMES e S.R. SUNSTEIN, *Il costo dei diritti. Perché la libertà dipende dalle tasse*, Bologna, 2000. Secondo gli autori non sono solo i diritti di welfare a dover essere finanziati dalle tasse, ma anche i cosiddetti "diritti negativi" (proprietà, libertà contrattuale e diritti della persona) che parrebbero realizzarsi per sola opera dei titolari. In realtà tali diritti non potrebbero essere esercitati senza la onerosa salvaguardia da parte dello stato attraverso forze dell'ordine, tribunali e servizi pubblici e servizi di ordine pubblico. Pertanto, la libertà individuale non sarebbe possibile senza un sistema fiscale che finanzia le istituzioni necessarie alla sua tutela.

³ Sulla funzione di tutela dei diritti sociali e di redistribuzione dei redditi da parte dei tributi importante è il contributo di F. GALLO, *Le ragioni del fisco*, Bologna, 2011.

⁴ Corte Costituzionale, sentenza 23 dicembre 2019, n. 288. Sul tema vedi anche L. ANTONINI (che della sentenza riportata è il redattore) *Dovere tributario, interesse fiscale e diritti costituzionale*, Padova, 1996. L'autore, pur riconoscendo rilevanza costituzionale all'interesse fiscale, non lo ritiene direttamente riconducibile all'articolo 2 della Costituzione e, pertanto, meritevole di una minore protezione rispetto ai diritti fondamentali del cittadino.

⁵ Interessante la diversa lettura che ne dà A. GIOVANNINI in *Note controvento su interesse fiscale e giustizia nell'imposizione come diritto fondamentale (muovendo dall'intelligenza artificiale)* in *La digitalizzazione dell'amministrazione finanziaria tra contrasto all'evasione e tutela dei diritti del contribuente*, Volume I, Milano, 2023. Secondo l'autore è errato affrontare il rapporto tra potere, interesse fiscale e diritti del contribuente affidandosi alla formula del bilanciamento, fondata sull'assunzione che il potere dello Stato si sostanzia nell'interesse fiscale e che quest'ultimo sia un principio costituzionale rilevante. Secondo Giovanni, infatti, i diritti fondamentali sono una dotazione giuridica propria dei loro titolari derivante direttamente dalla Costituzione. Non sono, pertanto, i diritti fondamentali a dover attendere l'esercizio del potere per il tramite della legge, ma è il potere, ad iniziare da quello legislativo, a doversi fermare innanzi ad essi.

⁶ Corte Costituzionale, sentenza 23 luglio 1991, n. 366 (ECLI:IT:COST:1991:366).

volte riconosciuto dalla giurisprudenza della Corte Costituzionale⁷ ed anche da quella civile, che per molti anni ne aveva negato l'esistenza, ammettendolo a partire dal 1975, riferendolo all'articolo 2 della Costituzione⁸, quale componente dei diritti della personalità. In ambito sovranazionale, la riservatezza trova protezione nell'articolo 8 (Diritto al rispetto della vita privata e familiare) della Convenzione Europea dei Diritti dell'Uomo e negli articoli 7 (Rispetto della vita privata e della vita familiare) e 8 (Protezione dei dati di carattere personale) della Carta dei diritti fondamentali dell'Unione Europea.

Il diritto alla riservatezza, in ambito tributario, trova un limite nell'interesse fiscale statale, oggetto di grandi approfondimenti in dottrina⁹ e giurisprudenza, che l'ha definito «non come uno degli interessi indistinti che sono affidati alla cura dell'amministrazione statale, ma come un interesse particolarmente differenziato, che attenendo al regolare funzionamento dei servizi necessari alla vita della comunità ne condiziona l'esistenza»¹⁰.

Si tratta, quindi, di effettuare un bilanciamento attraverso una tecnica che, individuato l'interesse prevalente, consenta di tutelarlo imponendo la minore compressione possibile per il diritto contrastante¹¹. Il principio cui ispirarsi è quello della proporzionalità¹² e implica la valutazione delle sue tre componenti: idoneità, necessità e proporzionalità in senso stretto. La verifica del requisito di idoneità comporta l'analisi della sussistenza del nesso di causalità tra mezzi e fini, ovvero l'adeguatezza degli strumenti a raggiungere il risultato. Il secondo passaggio del test di proporzionalità riguarda la necessità, ossia la valutazione se la misura utilizzata sia la meno lesiva per l'ottenimento del risultato in termini di sacrificio di altri diritti o interessi. La valutazione dell'ultima componente, la proporzionalità in senso stretto, comporta il raffronto tra i valori e beni giuridici in conflitto ed il grado della loro violazione. Il bilanciamento va valutato caso per caso poiché, come ha avuto modo di statuire al Corte costituzionale¹³, «tutti i diritti fondamentali tutelati dalla Costituzione si trovano in rapporto di integrazione reciproca e non è possibile, pertanto, individuare uno di essi che abbia la prevalenza assoluta sugli altri».

⁷ Corte Costituzionale, sentenze 21 febbraio 2019, n. 20 (ECLI:IT:COST:2019:20), 11 giugno 2009, n. 173 (ECLI:IT:COST:2009:173), 14 novembre 2006, n. 372 (ECLI:IT:COST:2006:372), 24 aprile 2002, n. 135 (ECLI:IT:COST:2002:135) e 11 marzo 1993, n. 81 (ECLI:IT:COST:2006:81).

⁸ Il *leading case* con il quale il diritto alla riservatezza è stato riconosciuto, dopo essere stato negato per molti anni (Cassazione sentenza 22.12.1956, n. 4487 (ECLI:IT:CASS:1956:4487)), è la sentenza Cassazione 27 maggio 1975, n. 2199 (ECLI:IT:CASS:1975:2199), con la quale la Corte ha ravvisato un diritto alla riservatezza nell'articolo 2 della Costituzione. Il caso riguardava la principessa Soraya Esfandiari rispetto alla quale la Corte ha affermato il diritto alla tutela della riservatezza per tutte le vicende strettamente personali che non hanno per i terzi interesse socialmente apprezzabile, anche se verificatesi fuori dal domicilio domestico.

⁹ Tra i molti si vedano P. MARONGIU, *L'amministrazione fiscale tra poteri e responsabilità*, Torino, 2016, 61; M. GREGGI, *Dall'interesse fiscale al principio di non discriminazione nella convenzione per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali*, in *Rivista di diritto finanziario*, 2001, 1, p. 412; A. MARCHESELLI, *Il giusto processo tributario in Italia. Il tramonto dell' "interesse fiscale"?*, in *Diritto e pratica tributaria*, 2001, I, p. 793.

¹⁰ Corte Costituzionale, sentenze 26 giugno 1965, n.50 (ECLI:IT:COST:1965:50); 6 giugno 1974, n. 164 (ECLI:IT:COST:1974:164).

¹¹ Per una approfondita analisi della tecnica del bilanciamento nell'ottica dei temi qui affrontati vedasi G. SCACCIA, «Proporzionalità e bilanciamento tra diritti nella giurisprudenza delle corti europee», in *Rivista AIC*, 2017, n. 3/2017, p. 1. Scettico sull'applicabilità del bilanciamento A. GIOVANNINI, *Note controvento*, cit.

¹² Per una disamina sulle origini del principio di proporzionalità vedasi D. U. GALLETTA, *Principio di proporzionalità e sindacato giurisdizionale nel diritto amministrativo*, Milano, 1998 e G. A. ANSALDI, *Principio di proporzionalità e funzioni pubbliche*. Acireale – Roma, 2012. Per l'applicazione del principio in ambito delle indagini e dell'accertamento tributario vedasi P. DALLE CARBONARE, *Principio di proporzionalità e sindacabilità giurisdizionale delle scelte istruttorie dell'Amministrazione Finanziaria in Il Commercialista Veneto*, n. 224, pag. 11.

¹³ Corte costituzionale, sentenza 9 maggio 2013, n.85 (ECLI:IT:COST:2013:85).

Il giudice delle leggi si è già cimentato nella valutazione del bilanciamento tra diritto alla riservatezza e interesse pubblico alla riscossione dei tributi¹⁴. Pronunciandosi in tema di segreto bancario, la Corte¹⁵ ha affermato che «il dovere di riservatezza connesso al segreto bancario non può coprire illeciti tributari e non può essere di ostacolo all'accertamento dei medesimi» poiché il dovere di riserbo cui sono tenute imprese bancarie non può costituire un «ostacolo all'adempimento di doveri inderogabili di solidarietà, primo fra tutti quello di concorrere alle spese pubbliche in ragione della propria capacità contributiva». Secondo la Corte, alla base del segreto bancario non ci sono valori della persona umana da tutelare ma semplicemente interessi economici e patrimoniali. Benché la verifica del diritto prevalente, come detto, debba essere effettuata con riferimento al caso concreto, la pronuncia in esame è piuttosto categorica nell'affermare la prevalenza del dovere di fedeltà fiscale rispetto a quello della riservatezza dei dati personali.

Anche in ambito europeo il diritto alla riservatezza dei dati non è un diritto assoluto, ma «va considerato alla luce della sua funzione sociale e va temperato con altri diritti fondamentali in ossequio al principio di proporzionalità», come ricorda il considerando n. 4 del Regolamento europeo in materia di protezione dei dati personali (GDPR). Sul tema la Corte di Giustizia ha affermato che la compressione del diritto alla riservatezza di un cittadino non contrasta con gli articoli 7 e 8 della Carta, poiché la riscossione delle imposte e la lotta all'evasione fiscale devono essere considerate compiti di interesse pubblico¹⁶ ed in virtù del fatto che tra le finalità dell'Unione vi è quella di combattere la frode e l'evasione fiscale internazionali¹⁷. È importante, tuttavia, che le deroghe e le restrizioni alla tutela dei dati personali siano rispettose del principio di proporzionalità, ossia intervengano entro i limiti dello stretto necessario¹⁸.

¹⁴ Per un'analisi della giurisprudenza domestica e della Corte di giustizia dell'Unione Europea, vedasi L. IZZO, *Il difficile rapporto tra il diritto alla privacy e il dovere di contribuzione alla spesa pubblica*, in *European Journal of Privacy Law & Technologies*, 2022, n. 2, p. 187.

¹⁵ Sentenza 18 febbraio 1992, n. 51. Il giudice delle leggi è stato chiamato a decidere sulla questione di legittimità degli articoli 63 del D.P.R. 633/1972 e 33 del D.P.R. 600/1973 sollevata dalla Commissione tributaria di Pordenone per eccesso di delega rispetto all'articolo 10, n. 12 della legge n. 825 del 1971, il quale conteneva la direttiva rivolta al legislatore delegato di stabilire «l'introduzione, limitata a ipotesi di particolare gravità, di deroghe al segreto bancario nei rapporti con l'amministrazione finanziaria, tassativamente determinate nel contenuto e nei presupposti». Secondo il giudice rimettente, con tale disposizione costituiva un riconoscimento del segreto bancario rispetto ad ogni attività ispettiva e di verifica posta in essere dall'amministrazione finanziaria: gli articoli predisposti dal legislatore delegato non prevederebbero la previsione di derogabilità al segreto bancario solo in casi limitati di particolare gravità.

¹⁶ Corte di Giustizia dell'Unione Europea, sentenza 27 settembre 2017, Puskas, C. 73/19 (ECLI:EU:C:2017:725). La Corte interpellata in merito alla legittimità di iscrizione da parte della direzione finanze di un cittadino in un registro pubblico delle persone considerate dei prestanome, ha ritenuto che ciò non contrasti con gli articoli 7 e 8 della Carta, poiché la riscossione delle imposte e la lotta all'evasione fiscale devono essere considerate compiti di interesse pubblico. La Corte ha delegato il giudice nazionale alla verifica della proporzionalità della misura, da valutarsi in termini di necessità per le finalità di contrasto all'evasione, minimizzazione della restrizione, idoneità al raggiungimento dello scopo ed inesistenza di mezzi meno restrittivi per raggiungerlo.

¹⁷ Corte di Giustizia dell'Unione Europea, sentenze 22 ottobre 2013, Sabou C-276/12, punto 32 (ECLI:EU:C:2013:678); 17 dicembre 2015, WebMindLicenses, C-419/14 (ECLI:EU:C:2015:832); 26 febbraio 2019 X, C-136/17 (ECLI:EU:C:2019:773).

¹⁸ Corte di Giustizia dell'Unione Europea, sentenze 21 dicembre 2016, Tele2 Sverige A B, C. 203/15 e 698/15, punto 96 (ECLI:EU:C:2016:970); 16 dicembre 2008, Satakunnan Markkinapörssi e Satamedia, C-73/07, punto 56 (ECLI:EU:C:2008:727); 9 novembre 2010, Volker und Markus Schecke e Eifert, C-92/09 e C-93/09, punto 77 (ECLI:EU:C:2010:662); 6 ottobre 2015, Schrems, C-362/14, Punto 92 (ECLI: EU:C:2015:650).

A livello sovranazionale, tuttavia, sia la Carta dei diritti fondamentali dell'Unione europea che la Convenzione europea per i diritti dell'uomo, prevedono una tutela specifica per il caso di ingerenza pubblica nella riservatezza del cittadino. Oltre al giudizio di bilanciamento in base al principio di proporzionalità, infatti, Convenzione e Carta condizionano la possibilità di limitazione del diritto alla riservatezza da parte di una pubblica autorità anche ad una specifica previsione da parte della legge¹⁹. Analoga prescrizione si ritrova nel Regolamento 679/2016 (GDPR) che, all'articolo 23²⁰ dispone che il diritto dell'Unione o dello Stato membro cui è soggetto il titolare o il responsabile del trattamento possano limitare, mediante misure legislative, la portata degli obblighi e dei diritti alla riservatezza, se la limitazione rispetta l'essenza dei diritti e delle libertà fondamentali e sia una misura necessaria e proporzionata in una società democratica.

La possibilità di comprimere il diritto alla riservatezza è, pertanto, legata alla duplice condizione della presenza di interessi collettivi che si pongano in condizione di superiorità nel giudizio di bilanciamento e di una legge che autorizzi la compressione del diritto. È il legislatore, in altri termini, che deve provvedere tassativamente a tipizzare i casi in cui l'amministrazione finanziaria è autorizzata alla raccolta ed alla conservazione dei dati dei contribuenti, determinandone la tipologia, i contenuti, i fini, le modalità di utilizzo. E non solo: la legge deve rispettare il contenuto essenziale dei diritti fondamentali: il mancato rispetto da parte di un atto dell'Unione comporta automaticamente la sua invalidità, senza che ci sia bisogno di valutare la proporzionalità della misura rispetto all'interesse generale perseguito²¹.

Sul tema non mancano interessanti spunti provenienti dalla Corte EDU, intervenuta a valutare la legittimità della legge ungherese che prevedeva la pubblicazione sul sito dell'autorità fiscale della lista dei grandi evasori. Con una prima sentenza la Corte ha ritenuto la norma proporzionata, rientrando nella disponibilità dello Stato le scelte in tema di politica sociale, economica e fiscale. La sentenza è stata ribaltata dalla Grande Camera, che ha ritenuto che la discrezionalità dello Stato in materia tributaria non sia illimitata, asserendo pertanto che la lista dei grandi evasori comportasse una violazione dell'articolo 8 della Convenzione. Con la sentenza la Corte ha posto anche i parametri con cui deve essere effettuato il bilanciamento, identificati nell'interesse pubblico alla divulgazione delle informazioni, nella natura e tipologia delle informazioni, nelle ripercussioni e rischi sulla vita privata dei contribuenti, nella tipologia del mezzo di comunicazione e nei principi fondamentali sulla protezione dei dati.

Si impone, tuttavia, una precisazione. Mentre l'articolo 8 della Cedu e l'articolo 7 della Carta prevedono che le limitazioni del diritto sia stabilita per legge, l'articolo 23 del GDPR

¹⁹ La formulazione delle disposizioni, sul tema, è dissimile. Nella Convenzione Europea dei Diritti dell'Uomo è riportata al comma 2 dell'articolo 8, il quale prevede che «non può esservi ingerenza della pubblica autorità nell'esercizio di tale diritto se non in quanto tale ingerenza sia prevista dalla legge e in quanto costituisca una misura che, in una società democratica, è necessaria per la sicurezza nazionale, l'ordine pubblico, il benessere economico del paese, la prevenzione dei reati, la protezione della salute o della morale, o la protezione dei diritti e delle libertà altrui». Nella Carta dei diritti fondamentali dell'Unione Europea il principio è enunciato dall'articolo 52, il quale impone che le limitazioni siano «previste dalla legge e rispettare il contenuto essenziale di detti diritti e libertà. Nel rispetto del principio di proporzionalità, possono essere apportate limitazioni solo laddove siano necessarie e rispondano effettivamente a finalità di interesse generale riconosciute dall'Unione o all'esigenza di proteggere i diritti e le libertà altrui».

²⁰ Regolamento UE 216/679 del Parlamento europeo e del Consiglio.

²¹ Corte di Giustizia dell'Unione europea, sentenza del 6 ottobre 2015, Schrems, C- 362/14 (ECLI: ECLI:EU:C:2015:650).

fa riferimento a “misure legislative”. A spiegare il significato di quest’ultimo termine soccorre il considerando 41 del regolamento 2016/679 (GDPR), il quale prevede che «qualora il presente regolamento faccia riferimento a una base giuridica o a una misura legislativa, ciò non richiede necessariamente l'adozione di un atto legislativo da parte di un parlamento, fatte salve le prescrizioni dell'ordinamento costituzionale dello Stato membro interessato». La Corte di Giustizia ha preso a riferimento tale indicazione, puntualizzando altresì che la base giuridica che consente l'ingerenza deve definire la portata della limitazione del diritto considerato²². La stessa Corte ha altresì affermato che tale normativa deve prevedere regole chiare e precise che fissino i requisiti minimi a garanzia della protezione dei dati contro rischi di abusi²³. Si deve tuttavia ritenere che tale interpretazione ampia del concetto di “misure legislative” non sia applicabile nel nostro ordinamento alla luce delle prescrizioni della Costituzione. In tema di limitazione dei diritti della libertà personale, del domicilio, della libertà e della segretezza della corrispondenza, la Carta prevede una duplice riserva, di legge e giurisdizionale. Se ne deve dedurre che anche la compressione del diritto alla riservatezza necessita di una legge approvata dal Parlamento²⁴.

Il legislatore domestico sembra invece pensarla diversamente, quantomeno in tema di riservatezza delle informazioni sui contribuenti. Già nel 2021²⁵ è stato modificato il Codice della Privacy con l'inserimento²⁶, tra le basi giuridiche legittimanti il trattamento dei dati personali, oltre alla legge, anche dei regolamenti e degli atti amministrativi generali e prevedendo²⁷ che il trattamento dei dati personali da parte di un'amministrazione pubblica sia consentito se necessario per l'adempimento di un compito svolto nel pubblico interesse o per l'esercizio di pubblici poteri ad esse attribuiti.

3. *La bulimica rincorsa alla raccolta di dati: un digiuno di diritti per il contribuente?*

Il reclutamento dell'intelligenza artificiale tra gli strumenti a disposizione dell'amministrazione finanziaria per le verifiche fiscali ha indotto molti studiosi ed operatori

²² Sul punto vedasi Corte di Giustizia sentenze 24 febbraio 2022, «SS» SIA, C. 175-2020, punto 52 e 54 (ECLI:EU:C:2022:124), 6 ottobre 2020, Privacy International, C-263/17, punto 65 (ECLI:EU:C:2020:790); sentenza del 16 luglio 2020, Facebook Ireland e Schrems, C-311/18 (ECLI:EU:C:2020:559); 17 dicembre 2015, WebMindLicenses, C-419/14, punto 81 (ECLI:EU:C:2015:832).

²³ Corte di Giustizia, sentenze 2 marzo 2021, Prokuratuur, C. 746/18 punto 48 (ECLI:EU:C:2021:152); 21 dicembre 2016, Tele2, C-203/15 e C-698/15, punti 117 e 118 (ECLI:EU:C:2016:970); 6 ottobre 2020, La Quadrature du Net e a., C-511/18, C-512/18 e C-520/18, punto 132 (ECLI:EU:C:2020:791).

²⁴ In senso conforme L. DEL FEDERICO, *Sui diritti dei contribuenti nell'era della trasformazione digitale*, in *Diritto e pratica tributaria internazionale*, n. 1, 1 gennaio 2024, p. 71 e A. CONTRINO, *Digitalizzazione dei poteri tributari e tutela dei diritti dei contribuenti: osservazioni generali sulle problematiche e sulle prospettive come introduzione al progetto di ricerca PRIN 2020*, in *La digitalizzazione dell'amministrazione finanziaria tra contrasto all'evasione e tutela dei diritti del contribuente*, volume 1, Milano, 2023, pag.19.

²⁵ D.L. 8 ottobre 2021, n. 139, art. 9, comma 1, lett. a), n. 1), convertito, con modificazioni, dalla L. 3 dicembre 2021, n. 205.

²⁶ D. Lgs. 30 giugno 2003, n.196, art. 2 ter, comma 1.

²⁷ D. Lgs. 30 giugno 2003, n.196, art. 2 ter, comma 1-bis.

del settore ad interrogarsi sulle insidie che il nuovo sistema potrà comportare alla riservatezza dei dati dei contribuenti²⁸.

A ben vedere, tuttavia, il tema della gestione delle informazioni non è nuovo, ma risale quantomeno all'epoca dell'istituzione dell'anagrafe tributaria²⁹: la riservatezza del contribuente diventa solo oggi d'attualità poiché, grazie all'impiego di potenti sistemi informatici, l'efficace interconnessione delle informazioni potrà consentire di estrarre indici di infedeltà fiscale a carico di soggetti o di schemi di operazioni.

Va, dunque, riformulata la domanda cui dare risposta, che non riguarda la legittimità dell'utilizzo dei sistemi di intelligenza artificiale in ambito tributario, quanto la legittimazione della raccolta, conservazione e trattamento delle informazioni contenute nelle banche dati a disposizione dell'amministrazione finanziaria.

Sotto questo profilo il nostro legislatore ha da tempo fatto la scelta di ritenere assolutamente prevalenti gli interessi dello Stato nell'esercizio del potere impositivo e del controllo fiscale rispetto ai diritti dei contribuenti alla tutela della riservatezza dei dati personali. Ne è testimonianza la straripante richiesta di informazioni da parte dell'Amministrazione finanziaria nei confronti di organi, istituzioni e contribuenti, in gran parte riversate nelle banche dati dell'anagrafe tributaria, che pone quella fiscale al vertice di tutte le amministrazioni in punto di disponibilità di dati.

La lettura della disposizione con la quale il Governo ha dato attuazione alla delega per la riforma tributaria³⁰ ne è ulteriore conferma. La lotta all'infedeltà fiscale potrà contare su una mole di dati davvero impressionante, a partire da quelli contenuti nell'anagrafe tributaria, la cui elencazione, almeno in questa sede, è praticamente impossibile: si va dai rapporti con operatori finanziari, ai dati autorizzativi in ambito edilizio, alle utenze di domestiche e non, ai pagamenti elettronici per interventi di recupero del patrimonio edilizio, agli estremi dei contratti conclusi dalle pubbliche amministrazioni e molti altri ancora. A ciò si aggiungano le informazioni extrafiscali acquisibili dalla fatturazione elettronica, in grado di fotografare abitudini, gusti, stile di vita dei cittadini, oltre ai prezzi e condizioni concessi da venditori e prestatori di servizi. Da ultimo, ma non per questo meno importanti, il legislatore della riforma tributaria ha incluso tra i dati utilizzabili per l'identificazione dei rischi di infedeltà fiscale anche tutte le informazioni pubblicamente disponibili, ossia quelle rilevabili dai social network o comunque liberamente consultabili in rete. La perplessità, in questo caso, riguarda le possibili interpretazioni da parte dei sistemi informatici di notizie non ufficiali, a volte artatamente distorte. Come interpreterà l'intelligenza artificiale una foto che ritrae un contribuente mentre siede orgoglioso a bordo di una macchina di gran lusso durante la visita ad una fiera motoristica? Probabilmente non lo sapremo mai, vista la mancanza di trasparenza offerta dalle machine learning in merito al procedimento che ha prodotto l'output.

Il Garante per la protezione dei dati personali ha più volte sottolineato³¹ che l'incetta di dati da parte dell'Amministrazione finanziaria non sia rispettosa dei principi di

²⁸ Incisivo A. GIOVANNETTI, *Note controvento*, cit. pag. 48, che afferma che «L'innesto delle nuove tecnologie su un terreno già arido di democraticità potrebbe determinare un ulteriore sbilanciamento dei termini del rapporto a danno dei diritti ed a favore del potere».

²⁹ Il fenomeno della protezione dei dati personali a disposizione dell'amministrazione finanziaria è stato sinora trascurato, come ricorda A. CONTRINO, *Banche dati tributarie scambio di informazioni fra autorità fiscali e "protezione dei dati personali": quali diritti e tutele per i contribuenti?* in *Rivista di diritto tributario (supplemento online)*, 29 maggio 2019, 1.

³⁰ D. Lgs. 12 febbraio 2024, n. 13, art. 2.

³¹ *Ex multis* vedasi Provvedimento 20 dicembre 2018 n. 511 con il quale il Garante raccomanda la minimizzazione della raccolta dei dati personali in materia di fatturazione elettronica. Sul tema vedi A.

proporzionalità e di minimizzazione della raccolta e conservazione dei dati. Non sempre, infatti, è possibile rilevare nelle richieste dell'amministrazione la necessità di ottenere i dati, visto che i risultati potrebbero essere raggiunti con altri mezzi che consentano un adeguato rapporto tra sacrificio imposto e beneficio raggiunto. Già in fase di osservazione del D.L. 26 ottobre 2019, n. 124, l'Authority³² aveva stigmatizzato l'eccesso di informazioni conservate e processate per la fatturazione elettronica, sottolineando i rischi derivanti dalla grande mole di informazioni trasmesse³³ riguardanti dettagli sui beni e i servizi ceduti, con la descrizione delle prestazioni, i rapporti fra cedente e cessionario e altri soggetti, riferiti anche a sconti applicati, fidelizzazioni, abitudini di consumo ed altri dati ancora, suggerendo al Governo di valutare l'effettiva necessità dell'integrale conservazione dei dati.

Lo sforzo di mettere a sistema i dati a disposizione dell'Amministrazione finanziaria mediante l'utilizzo di strumenti informatici e di intelligenza artificiale non è nuovo, e può essere rilevato nella sperimentazione dell'applicativo Ve.Ra.³⁴, volto ad esprimere liste selettive di contribuenti da sottoporre a verifica sulla base dell'interpolazione delle informazioni comunicate dagli operatori finanziari e delle altre informazioni esistenti nell'anagrafe tributaria. La verifica, secondo quanto previsto dal decreto attuativo, dovrebbe avvenire in base a dati preventivamente pseudonimizzati³⁵. Il Garante, in merito allo schema di provvedimento, ha avuto modo di affermare che le richieste di informazioni «possono ritenersi legittime solo qualora dall'esercizio dei diritti degli interessati che si intendono circoscrivere possa derivare un «pregiudizio effettivo e concreto» agli «interessi tutelati in materia tributaria e allo svolgimento delle attività di prevenzione e contrasto all'evasione fiscale, senza comprimarli in modo eccedente rispetto alle reali necessità perseguite»³⁶.

4. Sorveglianza algoritmica: il colpo di grazia ad un diritto agonizzante

È indubbio, comunque, che l'utilizzo dell'intelligenza artificiale non potrà che dare il colpo di grazia al diritto alla riservatezza dei dati del contribuente³⁷, già comunque agonizzante da una cinquantina d'anni. Non serve una sibilla per fare la previsione: giurisprudenza e legislatore sembrano aver da tempo decretato la supremazia dell'interesse

CONTRINO, *Protezione dei dati personali e pervasività delle banche dati fiscali in Digitalizzazione dell'amministrazione finanziaria tra contrasto all'evasione e tutela dei diritti del contribuente*, volume II, Milano 2023, pag. 110 e s.

³² Vedasi memoria Garante per la protezione dei dati personali del 5 novembre 2019.

³³ Nella memoria si parla di un flusso annuale di 2,1 miliardi di fatture.

³⁴ Ve.Ra. (Verifica dei rapporti finanziari) è un applicativo previsto dalla legge finanziaria per il 2020 (L. 27 dicembre 2019, n. 160 art. 1, comma 682) e attuato con il decreto MEF del 28 giugno 2022. Per un approfondimento G. CONSOLO, *Il nuovo applicativo di analisi del rischio fiscale "Ve.Ra" nel processo di digitalizzazione dell'attività dell'Amministrazione finanziaria*, in *Il diritto dell'economia* n. 112 (3 2023), pag. 153 e s.; D. CONTE, *Digitalizzazione, data protection e tecniche di profilazione nell'attività di accertamento tributario: quali diritti per i contribuenti?* in *La digitalizzazione dell'amministrazione finanziaria tra contrasto all'evasione e tutela dei diritti del contribuente*, volume II, Milano, 2023, pag. 133 e seg.

³⁵ La pseudonimizzazione è un procedimento di sostituzione delle anagrafiche tramite modifica delle variabili, in modo da impedire l'identificazione degli interessati. Sul tema vedi F. FARRI – *Digitalizzazione dell'amministrazione finanziaria e diritti dei contribuenti* in *Rivista diritto tributario*, 2020, pag. 125, nota 21.

³⁶ Garante per la protezione dei dati personali, Provvedimento 22 dicembre 2021, n. 453;

³⁷ Incisivo, ancora una volta, A. GIOVANNINI, *Note controvento*, cit., che afferma «(...) l'intelligenza artificiale potrebbe essere utilizzata come un cavallo di Troia per rinforzare l'interesse fiscale e piegare così i diritti del contribuente».

statale sul diritto alla riservatezza. Ne è testimonianza il decreto legislativo con il quale il Governo ha dato attuazione della delega per la riforma del diritto tributario: dopo aver previsto la facoltà di interconnessione tra le informazioni presenti nelle basi dati dell'Agenzia delle entrate per le attività di analisi del rischio fiscale e di controllo e stimolo dell'adempimento spontaneo, il decreto affida ad un emanando regolamento del Ministero dell'economia e delle finanze, sentito il Garante per la protezione dei dati personali, la definizione delle specifiche limitazioni e delle modalità di esercizio dei diritti afferenti la riservatezza del cittadino. Allo stesso regolamento è devoluta la definizione delle caratteristiche del trattamento previste dal citato articolo 23, paragrafo 2, del GDPR. Sarà dunque un regolamento del ministero cui fa capo l'ente deputato al controllo a determinare le modalità con le quali potrà essere limitato il diritto alla riservatezza del contribuente, con il solo onere di sentire il Garante. Un intervento, questo, che conferma la tendenza del legislatore domestico a considerare il diritto alla riservatezza del contribuente sempre recessivo rispetto all'interesse pubblico alla riscossione dei tributi.

La questione dell'utilizzo dell'intelligenza artificiale nel procedimento è stata sinora affrontata con particolare riguardo all'attività amministrativa. Il Consiglio di Stato³⁸, rifacendosi ad indicazioni di ordine sovranazionale, ha enunciato tre principi di cui tener conto ai fini della legittimità delle decisioni amministrative assunte attraverso strumenti informatici. Il primo è la conoscibilità da parte del diretto interessato dell'utilizzo di processi decisionali nei suoi confronti assunti con l'ausilio di algoritmi e a ricevere informazioni significative sulla logica utilizzata in tali processi. Il secondo principio riguarda la non esclusività della decisione algoritmica, ovvero il diritto che la decisione automatizzata sia controllata da una componente umana. Infine il principio, ripreso dal GDPR³⁹, che prevede il divieto di discriminazione algoritmica, inteso come approntamento di sistemi di controllo che evitino il trasferimento dei pregiudizi dell'uomo alla macchina ed inesattezze dei dati, minimizzando il rischio errori e massimizzando la sicurezza dei dati personali.

A livello normativo, l'Unione Europea per prima si è occupata dell'utilizzo dell'intelligenza artificiale, istituendo il Regolamento 2024/1689, comunemente denominato "AI Act". L'unico punto nel quale il regolamento si occupa dell'utilizzo dell'IA in campo tributario è il considerando 59, nel quale è previsto che «I sistemi di IA specificamente destinati a essere utilizzati per procedimenti amministrativi dalle autorità fiscali e doganali, come pure dalle unità di informazione finanziaria che svolgono compiti amministrativi di analisi delle informazioni conformemente al diritto dell'Unione in materia di antiriciclaggio, non dovrebbero essere classificati come sistemi di IA ad alto rischio utilizzati dalle autorità di contrasto a fini di prevenzione, accertamento, indagine e perseguimento di reati». Seppure sia stato anche diversamente interpretato⁴⁰, il considerando pare escludere dagli impieghi ad alto rischio i sistemi di intelligenza artificiale utilizzati dalle autorità fiscali e doganali solo quando utilizzati dall'autorità in attività che riguardano i reati.

I problemi indotti dall'utilizzo dell'intelligenza artificiale in ambito tributario, comunque, vanno ben oltre a quello della riservatezza dei dati del contribuente. Basti pensare

³⁸ Consiglio di Stato, sentenza 13 dicembre 2019, n. 8472 (ECLI:IT:CDS:2019:8472).

³⁹ Regolamento 679/2016, considerando 71.

⁴⁰ M. CEDRO, A. CIMINO, *Intelligenza artificiale ed accertamento tributario: opportunità e rischi di una rivoluzione in atto in Diritto e pratica tributaria internazionale*, n. 1., gennaio 2024, pag. 48, D. CONTE, *Digitalizzazione*, cit., pag. 159, L. DEL FEDERICO, *Sui diritti dei contribuenti*, cit. p.72. Gli autori, che commentano una versione non definitiva nella quale il considerando era posto al numero 38 con diversa formulazione, ritengono che l'esclusione riguardi tutti i sistemi di AI utilizzati nei procedimenti amministrativi tributari.

alla mancanza di trasparenza degli algoritmi, che rende gli output utilizzabili solo per la scelta dei contribuenti da assoggettare a controllo, ma non quale ausilio per le attività di accertamento, nelle quali è necessario fornire la motivazione e gli elementi di prova a supporto, tutti elementi non ottenibili, allo stato, dell'intelligenza artificiale⁴¹.

I rischi derivanti dall'utilizzo dell'algoritmo nel procedimento di indagine e accertamento tributari sono, al momento, meramente potenziali. Si deve, infatti, prendere atto che il decreto attuativo della delega fiscale prevede l'utilizzo dei sistemi di intelligenza artificiale limitata alla fase di analisi del rischio, con finalità di prevenzione e contrasto all'evasione fiscale, alla frode fiscale e all'abuso del diritto in materia tributaria, nonché di stimolo dell'adempimento spontaneo, e lo svolgimento di controlli preventivi⁴². Non sembra, allo stato, che tale previsione possa ricomprendere l'utilizzo dell'algoritmo anche alla predisposizione degli atti di accertamento sul singolo contribuente. Va, piuttosto, segnalato che la compressione del diritto alla riservatezza non si applica solo per il contrasto dell'infedeltà fiscale, ma anche per la prevenzione. Il sacrificio della riservatezza dei dati, in altri termini, è previsto non solo per i soggetti sospettati di comportamenti evasivi, ma indiscriminatamente per tutti i contribuenti.

Si pone, a questo punto, una legittima domanda: resisterà il nostro legislatore alla tentazione di promuovere il risultato dell'intelligenza artificiale al rango di presunzione legale relativa? I precedenti ci sono, basti pensare agli studi di settore, a lungo ritenuti idonei a integrare presunzioni da sole sufficienti ad assicurare un valido fondamento all'accertamento tributario, salva la possibilità, per il contribuente, di fornire la prova contraria. Valida testimonianza di tale tendenza è il concordato preventivo biennale⁴³, che consente di predeterminare il reddito imponibile dei due anni successivi sulla base di un'elaborazione algoritmica dei dati forniti dal contribuente con gli indicatori sintetici di affidabilità. In questo caso, l'output proposto dalla macchina non può essere messo in discussione dal contribuente, che può solo decidere se aderire o meno al concordato. Si tratta di uno strumento che, probabilmente, preannuncia una nuova era del rapporto tra contribuente e fisco, ben lontano dai principi di contraddittorio e di obbligo di motivazione degli atti che sono stati riaffermati proprio con l'attuazione della legge delega.

5. Conclusioni

Il dovere tributario, nella sua più alta accezione costituzionale definita in apertura del presente lavoro, rappresenta un pilastro del patto sociale: non una mera obbligazione giuridica, ma l'espressione concreta della solidarietà dei cittadini, funzionale alla realizzazione dei diritti fondamentali ed all'equità redistributiva. Tuttavia, nel passaggio dal piano dei principi a quello pratico, si assiste una progressiva distorsione di questo disegno, che vede prevalere l'interesse fiscale a scapito della salvaguardia dei diritti inviolabili del contribuente,

⁴¹ In senso conforme S. DORIGO, "Opportunità e limiti all'impiego dell'intelligenza artificiale da parte del Fisco" in *Corriere Tributario* n. 11 2022 pag. 965.

⁴² Il D.Lgs. 12 febbraio 2024, n. 13, art. 2, prevede l'utilizzo delle interconnessioni tra le banche dati esclusivamente «al fine di massimizzare l'efficacia delle attività di prevenzione e contrasto all'evasione fiscale, alla frode fiscale e all'abuso del diritto in materia tributaria, nonché di quelle volte a stimolare l'adempimento spontaneo».

⁴³ Previsto dalla legge delega 9 agosto 2023, n. 111 ed attuato dal D.Lgs. 13 dicembre 2023, n. 221.

primo fra tutti quello della riservatezza. L'avvento dell'intelligenza artificiale non potrà che accentuare tale squilibrio, caratterizzato dal controllo indiscriminato, collettivo e preventivo sul contribuente. I possibili sviluppi di questo scenario fanno prevedere non una semplice erosione del diritto alla privacy, ma una sua strutturale marginalizzazione nel nome dell'efficienza e della lotta all'infedeltà fiscale, con una progressiva "algoritmizzazione" del rapporto tra contribuente e amministrazione e delle funzioni di indagine ed accertamento in particolare.