

Intelligenza artificiale generativa e giustizia: profili costituzionali tra efficienza e tutela dei diritti fondamentali

di *Erik Longo*

*La legge non è una macchina e i giudici
non sono dei macchinisti¹*

Questo articolo indaga alcune delle molteplici intersezioni tra intelligenza artificiale (IA) e diritto, concentrandosi sulle implicazioni dell'uso delle più recenti forme di IA generativa nei processi giudiziari. La prima parte è dedicata a un'analisi del dibattito precedente al 2022. Nella seconda parte, dopo aver introdotto la sottocategoria dei *Large Language Models*, si esamina l'uso amatoriale di ChatGPT in alcuni casi recenti, evidenziando i potenziali benefici, come il miglioramento dell'efficienza e dell'efficacia delle decisioni, e i rischi relativi alla trasparenza, alla protezione dei dati personali, all'indipendenza giudiziaria e al diritto a un processo equo. L'articolo esplora poi alcuni documenti istituzionali che regolano lo sviluppo della giustizia digitale in diverse giurisdizioni e affronta il tema del processo decisionale umano assistito dall'IA. Nell'ultima parte si discute se siamo all'alba di una nuova forma di giustizia, sottolineando l'esigenza di bilanciare i progressi tecnologici con la tutela dei diritti fondamentali.

1. Introduzione / 1.1. *L'uso dell'IA nel processo deve rispettare le dimensioni della giustizia* / 1.2. *La rilevanza del tema* / 1.3. *La prospettiva costituzionalistica* / **2.** Gli anni dieci: l'impatto con i problemi del *Machine Learning* / 2.1. *L'uso "ausiliario" dell'IA e i problemi derivanti dallo "human in the loop"* / 2.2. *Due rischi chiave nell'uso assistivo dell'IA nel processo* / 2.3. *Bias, opacità e diritto antidiscriminatorio* / **3.** Gli anni venti: l'arrivo dell'IA generativa (*GenAI*) / 3.1. *Gli usi dei LLM nel campo del diritto, con una prima impressione su di essi* / 3.2. *Dai problemi tecnici alle ricadute costituzionali* / 3.3. *Le strategie per risolvere tali problemi* / **4.** *La disciplina europea e italiana sull'IA nel settore giustizia* / 4.1. *La "giustizia" nell'AI Act* / 4.2. *La legge n. 132/2025* / **5.** *I documenti di indirizzo circa l'uso della GenAI nei processi* / 5.1. *Le raccomandazioni del Csm* / 5.2. *Le altre linee guida a livello internazionale* / **6.** *Profili costituzionali relativi all'utilizzo dei LLM nei processi* / **7.** *Considerazioni conclusive*

35

35

1. Introduzione

I sistemi di Intelligenza Artificiale hanno compiuto un percorso straordinario, passando da innovazioni tecnologiche di nicchia a strumenti sempre più radicati nella pratica giuridica a livello globale². Lungi dall'annunciare il tramonto del giudice umano, la crescente integrazione dell'IA nelle funzioni giudiziarie sta ridefinendo in profondità la natura stessa del lavoro fuori e dentro le aule giudiziarie, trasformando i compiti professionali, riorganizzando le strutture istituzionali e rimodellando il valore attribuito alla competenza giuridica³. Molte amministrazioni giudiziarie se ne servono per migliorare l'accesso alle informazioni, ottimizzare la gestione dei casi e supportare il processo decisionale dei giudici⁴.

¹ J. FRANK, *Law and the modern mind*, Transaction Publishers, London, 1930.

² B. ALARIE, e al., *Law in the Future*, in *University of Toronto Law Journal*, 4, 2016, 423–428.

³ T.V.H. NGUYEN, *Artificial Intelligence and the Transformation of Legal Practice: From Automation to Augmented Lawyering*, in *Management Research Quarterly*, 3, 2026, 19–27.

⁴ S.D. HODGE JR, *Revolutionizing justice: unleashing the power of artificial intelligence*, in *SMU Sci. & Tech. L. Rev.*, 2, 2023, 217–258

In questa prospettiva, l'IA rappresenta un potente strumento per ampliare l'accesso alla giustizia e promuovere decisioni più eque ed efficienti⁵. Ma fino a che punto si può spingere il suo impiego nei processi? La domanda non è puramente teorica. Se oggi l'IA opera ancora come strumento di supporto – subordinato alla supervisione umana – la possibilità che in futuro il ragionamento giudiziario venga affidato a sistemi automatici non è così remota⁶. Le norme europee vigenti sembrano offrire garanzie sufficienti contro questo scenario⁷, ma il dibattito normativo e istituzionale sul punto non può ancora considerarsi concluso.

Accettare che un algoritmo possa sostituire il giudizio di un essere umano significa ammettere che decisioni che incidono profondamente sulla vita delle persone vengano sottratte a qualsiasi forma di responsabilità etica e revisione critica⁸. È dunque lecito chiedersi: cosa ci garantisce davvero che la supervisione umana non venga progressivamente erosa, fino a diventare una formalità, di fronte a tecnologie che ogni giorno diventano via via più performanti⁹?

Va chiarito che la preoccupazione di fondo da cui derivano tali domande si estende alla dinamica futura tra esseri umani e macchine¹⁰. Il pericolo imminente non è la cessione del controllo alle macchine e alla tecnologia, bensì l'emergere di un nuovo tipo di autorità che potrebbe emarginare gli individui e limitarne i diritti¹¹.

1.1. L'uso dell'IA nel processo deve rispettare le dimensioni della giustizia

Alcuni anni fa, il Consiglio d'Europa ha precisato che l'introduzione dell'IA nel processo deve necessariamente essere limitata e governata da una serie di principi che disciplinano l'impatto di tale sistema tecnico nelle diverse dimensioni della giustizia¹². Ancor prima, i

⁵ Si v. a tale proposito l'imponente progetto di documentazione su questi temi realizzato dall'Oxford Institute of Technology and Justice e qui reperibile: <https://www.techandjustice.bsg.ox.ac.uk/>. Va rilevato che il tema non è nato ora ma che ha una storia oramai ultracinquantennale. Risalgono ai primi anni '70 del secolo scorso i primi contributi che hanno applicato tecniche di intelligenza artificiale al diritto. Allora si trattava dei cosiddetti "sistemi esperti" usati al fine di ottenere indicazioni sulla soluzione giuridica di casi concreti. La dottrina sul tema è pressoché sterminata. *Infra multis* v. M.G. LOSANO, *Giuscibernetica: macchine e modelli cibernetici nel diritto*, Einaudi, Torino, 1969; B.G. BUCHANAN, T.E. HEADRICK, *Some speculation about artificial intelligence and legal reasoning*, in *Stan. L. Rev.*, 1, 1970, 40–62; V. FROSINI, *La giuritecnica: problemi e proposte*, in *Informatica e diritto*, 1, 1975, 26–35; A. D'AMATO, *Can/should computers replace judges*, in *Ga. L. Rev.*, 5, 1976, 1277–1301; L.T. MCCARTY, *Reflections on TAXMAN: An experiment in artificial intelligence and legal reasoning*, in *Harv. L. Rev.*, 5, 1976, 837–893; R.E. SUSSKIND, *Expert systems in law: A jurisprudential approach to artificial intelligence and legal reasoning*, in *Mod. L. Rev.*, 2, 1986, 168–194.

⁶ A. SIMONCINI, *La dimensione costituzionale della giustizia predittiva. Riflessioni su intelligenza artificiale e processo*, in *Riv. dir. proc.*, 2, 2024, 389–423.

⁷ Si v. a questo proposito il testo del Regolamento UE n. 2024/1689, c.d. AI Act.

⁸ M.G. CIVININI, *The Perks of Using AI Tools in the Justice Domain*, in *Artificial Intelligence, Judicial Decision-Making and Fundamental Rights*, a cura di Scuola Superiore della Magistratura, https://www.julia-project.eu/sites/default/files/2025-02/JuLIA_handbook_Justice_final.pdf, 2024, pp. 20 ss.

⁹ F. CONTINI, *Unboxing Generative AI for the Legal Professions: Functions, Impacts and Governance*, in *International Journal for Court Administration*, 2, 2024, 1–22.

¹⁰ S.E. BIBER, *Between Humans and Machines: Judicial Interpretation of the Automated Decision-Making Practices in the EU*, 2023, University of Luxembourg Law Research Paper (reperibile in: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4662152)

¹¹ A. SIMONCINI, E. LONGO, *Digital Constitutionalism and Fundamental Rights: Reconfiguring Liberty and Power*, in *The Oxford Handbook of Digital Constitutionalism*, a cura di G. DE GREGORIO, et al., Oxford University Press, Oxford, 2026, p. 25 ss.

¹² Come già aveva indicato la Commissione europea per l'efficienza della giustizia del Consiglio d'Europa (CEPEJ) il 3-4 dicembre 2018, a Strasburgo, approvando il documento "European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment". Il testo in italiano è qui reperibile: CEPEJ, *Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*, 2018 (reperibile in: <https://rm.coe.int/carta-etica-europea-sull-utilizzo-dell-intelligenza-artificiale-nei-si/1680993348>).

filosofi più avveduti avevano avvertito che nessun sistema tecnico può manifestarsi, svilupparsi ed esistere se non inserendosi in un corpo sociale esistente, il quale viene da esso modificato (in una metamorfosi reciproca tra lo stesso sistema tecnico e il corpo sociale relativo)¹³.

L'IA applicata al processo, quindi, non solo non può produrre “ingorghi e nocività”, ma deve rispettare la società e l’“ambiente umano e «naturale»¹⁴” in cui è collocata, anche al di là della sua enunciazione legale¹⁵; è qui che emerge il supremo principio secondo cui la decisione giudiziaria deve essere sempre presa dall'essere umano¹⁶.

Ma se occorre mantenere l'essere umano in capo a questi usi dell'IA, bisogna pure assicurare un quadro composito di principi, valori e regole che permettano di “domare” la tecnica secondo le necessità umane, agendo direttamente sulle conseguenze che essa comporta¹⁷. Dopo la sperimentazione, infatti, occorrono norme che consentano di creare un quadro di governance adeguato, capace di bilanciare innovazione e tutela dei diritti, assicurando che l'avanzamento tecnologico serva ai valori democratici e contribuisca allo sviluppo della società intera¹⁸.

Nei fatti, oggi molte corti e tribunali nel mondo sono interessati ad adottare strumenti di IA a supporto di una serie di attività giudiziarie¹⁹. Ognuna di queste applicazioni dell'IA nel settore giudiziario è stata promossa per migliorare l'efficienza, l'efficacia e l'accesso alla giustizia²⁰. Non sempre, però, gli effetti sono stati quelli sperati, sia per via di problemi tecnici sia di problemi giuridico-sociali.

¹³ L'inserimento della tecnica in quel contesto non lascia intatto quel corpo sociale, né consente ai suoi diversi elementi di svilupparsi in modo autonomo e indipendente dall'essere umano. J. ELLUL, *Il sistema tecnico. La gabbia delle società contemporanee*, Prefazione di Jean-Luc Porquet, Jaca book, Milano, 2009, p. 107.

¹⁴ Op. ult. cit., p. 149.

¹⁵ Una delle espressioni più interessanti di questo principio si trova nel Considerando n. 61 del Regolamento (Ue) 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale (AI Act), dove si legge: “l'utilizzo di strumenti di IA 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”.

¹⁶ In dottrina il principio dello “human oversight” è stato indicato con forza e autorevolezza da G. GALLONE, *Riserva di umanità e funzioni amministrative*, CEDAM - Wolters Kluwer, Milano, 2023.

¹⁷ Ellul ci ricorda che la tecnica non è come i fenomeni naturali, che sono capaci di “autoregolarsi”. J. ELLUL, *Il sistema tecnico. La gabbia delle società contemporanee*, cit., p. 150.

¹⁸ Come ricorda, in un interessante articolo recente, M.G. LOSANO, *Le Corti costituzionali di Romania e Colombia sulle interferenze informatiche e sull'intelligenza artificiale*, in *Diritto pubblico comparato ed europeo*, 1, 2025, 117–138.

¹⁹ Gli usi testati sono i seguenti: per la ricerca, la revisione e l'individuazione di documenti su larga scala; per l'aiuto alla risoluzione automatizzata delle controversie online; per la previsione di eventi futuri (ad esempio, l'esito di contenziosi); per l'anonimizzazione e la pseudonimizzazione delle decisioni giudiziarie; per la classificazione dei documenti archiviati; per la trascrizione (voce-testo) e la traduzione (sia voce che testo). Di usi dell'IA nei processi si discute da moltissimo tempo, conseguentemente esiste una letteratura sconfinata. *Infra multis*, si v.no: R.E. SUSSKIND, *Expert systems in law: A jurisprudential approach to artificial intelligence and legal reasoning*, cit.; C. TATA, *The Application of Judicial Intelligence and Rules to Systems Supporting Discretionary Judicial Decision-Making*, in *Artif. Intell. Law*, 2-4, 1998, 203–230; K.D. ASHLEY, S. BRÜNINGHAUS, *Automatically classifying case texts and predicting outcomes*, in *Artificial Intelligence and Law*, 2, 2009, 125–165; D. REILING, *Technology for justice: How information technology can support judicial reform*, Leiden University Press, Leiden, 2009; G. LUPO, *Regulating (Artificial) Intelligence in Justice: How Normative Frameworks Protect Citizens from the Risks Related to AI Use in the Judiciary*, in *European Quarterly of Political Attitudes and Mentalities*, 2, 2019, 75–96; G. CONTISSA, G. LASAGNI, *When it is (also) Algorithms and AI that decide on Criminal Matters: In Search of an Effective Remedy*, in *European Journal of Crime, Criminal Law and Criminal Justice*, 3, 2020, 280–304, 280; S. QUATTROCOLO, *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, Springer, Cham, 2020; D. REILING, *Digital Justice: Nice to Have but Hard to Achieve*, in *Technology, Innovation and Access to Justice*, a cura di S.P. DE SOUZA, M. SPOHR, Edinburgh University Press, Edinburgh, 2022, 131–143, 131.

²⁰ N. KYRIAKIDES, e al., *AI and access to justice: An expansion of Adrian Zuckermans findings*, in *Frontiers in Civil Justice*, a cura di X. KRAMER, et al., Edward Elgar Publishing, Cheltenham, 2022, 121–141; X.E. KRAMER, *Access to Justice and Technology: Transforming the Face of Cross-Border Civil Litigation and Adjudication in the EU*, in *eAccess to Justice*, a cura di K. BENYEKHELF, et al., University of Ottawa Press, Ottawa, 2017, 351–376; C.V. CHIEN,

Nei tempi più recenti, i giudici, i pubblici ministeri, il personale di cancelleria e gli avvocati hanno iniziato a utilizzare chatbot basati su modelli linguistici di grandi dimensioni (LLM) – spesso senza autorizzazione – per redigere documenti legali e decisioni giudiziarie ed elaborare documenti per le udienze²¹. Alcuni tribunali in giro per il mondo hanno iniziato anche a fornire ai propri magistrati l'uso di sistemi di IA appositamente sviluppati per aiutarli nella ricerca della giurisprudenza. Altri hanno fornito licenze per alcuni LLM specifici per aiutare nella redazione di atti.

Insieme alle sperimentazioni e alla fornitura di appositi tool, sono nate iniziative locali o internazionali rivolte allo scopo di creare specifiche *policy* per evitare l'uso improprio di tali strumenti²².

Non mancano, purtroppo, gli esempi di usi distorti dei sistemi di IA, per lo più in controversie in cui giudici o avvocati riportano in atti ufficiali citazioni false generate dall'interrogazione di un LLM. Alcuni di questi casi si sono conclusi addirittura con un procedimento disciplinare o con l'irrogazione di una multa²³.

Forse il caso più eclatante è quello accaduto alcuni anni fa in Colombia nel corso di una causa sull'attribuzione di una prestazione sociale legata al diritto alla salute. All'interno di quella vicenda, nell'agosto 2024, la Corte costituzionale dello Stato sudamericano ha emesso una sentenza (n. T-323/2024) che esamina una questione di costituzionalità legata a una *acción de tutela* così riassumibile: “rispetta il diritto al giusto processo un giudice che nel decidere l'aggiudicazione del diritto alla salute incorpora nella sua sentenza la richiesta di risolvere il caso rivolta a ChatGPT?”²⁴.

Proprio a partire dalla questione di costituzionalità sollevata nell'“azione”, i giudici costituzionali colombiani hanno stabilito che l'uso dell'IA da parte della magistratura è “costituzionale” nella misura in cui la macchina non sostituisca il giudizio umano, in particolare nelle decisioni giudiziarie²⁵. I giudici costituzionali hanno statuito che i sistemi di IA possono essere utilizzati dalla magistratura per supportare compiti gestionali (ad esempio, prendere appunti e redigere relazioni), compiti documentali (ad esempio, la classificazione dei documenti), funzioni giudiziarie (ad esempio, l'identificazione della giurisprudenza pertinente) e per correggere e sintetizzare testi sotto la supervisione dell'utente²⁶.

M. KIM, *How Generative AI Can Help Address the Access to Justice Gap Through the Courts*, 2024, (reperibile in: <https://ssrn.com/abstract=4683309>).

²¹ Sia consentito rimandare a E. LONGO, *Giustizia digitale e Costituzione: riflessioni sulla trasformazione tecnica della funzione giurisdizionale*, FrancoAngeli, Milano, 2023.

²² UNESCO, *Guidelines for the Use of AI Systems in Courts and Tribunals*, (reperibile in: <https://www.unesco.org/en/articles/guidelines-use-ai-systems-courts-and-tribunals>).

²³ I casi censiti sono attualmente poco meno di mille. Si v. il ricco database disponibile all'indirizzo: <https://www.damiencharlotin.com/hallucinations/>

²⁴ Il contenuto creato dal chatbot consisteva in risposte a quattro domande giuridiche che il giudice aveva trascritto come parte della motivazione della sua sentenza. Sul caso si v. J.D. GUTIÉRREZ, *ChatGPT in Colombian Courts: Why we need to have a conversation about the digital literacy of the judiciary*, VerfBlog, 23/2/2023 (reperibile in: <https://verfassungsblog.de/colombian-chatgpt/>); R. PERONA, *Giustizia artificiale intelligente? Spunti per il costituzionalista europeo a partire dall'esperienza colombiana*, in *Federalismi.it*, 26, 2025, 98–129

²⁵ In general on these issues, see J. MORISON, A. HARKENS, *Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making*, in *Legal Studies*, 4, 2019, 618–635.

²⁶ La Corte costituzionale colombiana è passata poi a delineare una serie di condizioni costituzionali da garantire qualora i membri della magistratura desiderino utilizzare sistemi di IA a supporto delle loro attività giudiziarie: (i) trasparenza, (ii) responsabilità, (iii) privacy, (iv) non sostituzione della razionalità umana, (v) serietà e verifica, (vi) prevenzione dei rischi, (vii) uguaglianza ed equità, (viii) controllo umano, (ix) regolamentazione etica, (x) conformità alle buone pratiche e agli standard collettivi, (xi) monitoraggio e adattamento continui e (xii) adeguatezza. A commento di questi aspetti della sentenza si v. B.A. HERRERA-TAPIAS, e al., *Algorithmic discrimination and explainable artificial intelligence in the judiciary: A case study of the Constitutional Court of Colombia*, in *Procedia Computer Science*, 2025, 1227–1232; R. PERONA, Y.C. DE LA ROSA, *Unveiling AI in the courtroom: exploring ChatGPT's impact on*

1.2. La rilevanza del tema

Il momento attuale considera l'uso dell'IA nella giustizia tra gli aspetti più intricati dell'intera trasformazione della giustizia. Anche la legislazione europea tiene conto della rischiosità insita negli usi dell'IA nell'ambito della giustizia. L'EU AI Act classifica come “ad alto rischio” quei sistemi di IA che sono “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 controversie²⁷”. Tale classificazione comporta una serie di obblighi per i responsabili dell'implementazione di questi sistemi, quali l'attuazione di un sistema di gestione dei rischi e la garanzia di un controllo umano sugli strumenti²⁸.

La legge italiana n. 132/2025 prevede all'art. 15 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” e che spettano al Ministero della Giustizia sia disciplinare gli impieghi dell'IA per l'organizzazione e la semplificazione degli uffici giudiziari sia, in attesa dell'attuazione dell'AI Act stesso, la “sperimentazione e l'impiego dei sistemi di intelligenza artificiale negli uffici giudiziari”²⁹.

La regolamentazione del fenomeno esprime quanto già detto in precedenza citando Ellul: l'integrazione della tecnica in qualsiasi ambito sociale richiede regole in grado di definire l'assetto dei poteri e delle responsabilità. Nel caso del sistema giudiziario, ciò diviene ancora più importante, giacché si toccano da vicino quei fili invisibili, ma molto delicati, che salvaguardano la separazione dei poteri, la tutela dei diritti e, da ultimo, l'esistenza di un ordinamento democratico³⁰.

È per queste ragioni che qualsiasi uso di un sistema di IA non dovrebbe mai ostacolare la competenza, l'indipendenza e l'imparzialità della magistratura perché, come affermato nei considerando dei “Principi di condotta giudiziaria di Bangalore”³¹, tali attributi sono essenziali per la tutela dei diritti umani e perché “l'attuazione di tutti gli altri diritti dipende in ultima analisi dalla corretta amministrazione della giustizia”³².

Inoltre, se è vero che l'impiego dei sistemi di IA per l'organizzazione dei tribunali può garantire il diritto a ragionevoli tempi processuali, è noto anche che la loro implementazione potrebbe contribuire a discriminare sistematicamente individui o gruppi vulnerabili o emarginati, o a limitare l'accesso all'ecosistema digitale e ai relativi dispositivi³³.

judicial decision-making through a pilot Colombian case study, in *AI & SOCIETY*, 4, 2025, 2533–2540; R. PERONA, *Giustizia artificiale intelligente? Spunti per il costituzionalista europeo a partire dall'esperienza colombiana*, cit.

²⁷ V. Allegato n. 3 dell'AI Act.

²⁸ A. PUNZI, *La decisione giudiziaria nell'AI Act*, in *Giur. it.*, 2, 2025, 448–452.

²⁹ V. art. 15, comma 3. Il comma 4 dell'art. 15 attribuisce inoltre al Ministero della Giustizia il compito di curare la formazione dei magistrati e del personale dei tribunali.

³⁰ F. DAL CANTO, *Lezioni di ordinamento giudiziario*, II ed., Giappichelli, Torino, 2020; N. ZANON, F. BIONDI, *Il sistema costituzionale della magistratura*, V ed., Zanichelli, Bologna, 2019.

³¹ I Principi di Condotta Giudiziaria di Bangalore (The Bangalore Principles of Judicial Conduct) sono stati adottati dalle Nazioni Unite nel 2002 all'esito del lavoro di un gruppo composto da alti rappresentanti delle magistrature di Paesi dei diversi continenti, con l'apporto dei componenti del Consiglio consultivo dei giudici europei (CCJE). V. https://www.unodc.org/res/ji/import/international_standards/bangalore_principles/bangalore_principles_italian.pdf

³² Ciò è particolarmente rilevante se si guarda per esempio all'importanza degli artt. 24 e 113 della Costituzione italiana. G. SORRENTI, *Giustizia e processo nell'ordinamento costituzionale*, Giuffrè, Milano, 2013.

³³ C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi*, in *DPCE Online*, 3, 2020, 3369–3389.

Prima di creare problemi tecnici e possibilità di “singolarità giuridica”³⁴, l’uso di tecnologie digitali avanzate in campo giuridico solleva questioni relative alla tutela dei diritti fondamentali e alla legittimità di soluzioni tecnologiche che tendono a catturare i decisori e a trasformare le loro attività in meri processi matematici o statistici³⁵.

1.3. La prospettiva costituzionalistica

In questo ampio orizzonte, anche la prospettiva del diritto costituzionale può costituire un utile punto di vista³⁶. Questa disciplina ha il duplice scopo di analizzare e tutelare i diritti e di definire i limiti legittimi ai poteri delle istituzioni pubbliche³⁷.

Negli ultimi anni il diritto costituzionale ha contribuito a definire i contorni generali del rapporto tra diritto e tecnologia, in particolare in relazione ai profili della tutela dei diritti e al rapporto tra i poteri³⁸. I risultati delle analisi dottrinarie hanno aiutato a comprendere che l’interazione tra il digitale e il diritto non è soltanto una questione tecnica, ma un insieme di problemi che possono essere affrontati con gli strumenti a disposizione di ogni giurista, come dimostrano le analisi di coloro che si occupano dell’impatto del processo decisionale automatizzato sui diritti fondamentali³⁹. Gli studiosi di diritto costituzionale hanno analizzato ampiamente le strategie per rendere le decisioni automatizzate più coerenti ed eque, riducendo il rischio di errori umani, di violazione del principio di legalità e di abuso di potere⁴⁰.

Tali analisi evidenziano, da un lato, che la legge da sola non basta a impedire la formazione di pregiudizi ‘algoritmici’ e, dall’altro, che l’impiego dell’IA comporta rischi significativi per i diritti e le libertà individuali, soprattutto quando riproduce i pregiudizi già presenti nei dati di addestramento e formula previsioni approssimative⁴¹. Queste pratiche possono portare a risultati indesiderati e ingiustificati, come discriminazioni e violazioni della libertà di scelta individuale, negando, in ultima analisi, l’accesso a beni o servizi⁴².

³⁴ J. COBBE, *Legal Singularity and the Reflexivity of Law*, in *Is Law Computable? Critical Perspectives on Law and Artificial Intelligence*, a cura di S. DEAKIN, C. MARKOU, Hart, London, 2020, , 107.

³⁵ A. SIMONCINI, *Il linguaggio dell’Intelligenza Artificiale e la tutela costituzionale dei diritti*, in *Rivista AIC*, 2, 2023, 1–39.

³⁶ A. SIMONCINI, E. LONGO, *Fundamental Rights and the Rule of Law in the Algorithmic Society*, in *Constitutional Challenges in the Algorithmic Society*, a cura di H.-W. MICKLITZ, et al., Cambridge University Press, Cambridge, 2022, 27–40, 27.

³⁷ Su questo tema si veda, in generale, C.H. MCILWAIN, *Constitutionalism: Ancient and Modern*, Revised edition, Cornell University Press, Ithaca, NY, 2005; M. FIORAVANTI, *Constitutionalism*, in *A Treatise of Legal Philosophy and General Jurisprudence*, a cura di D. CANALE, et al., 9, Springer, 2009, 263–300; N.W. BARBER, *The Principles of Constitutionalism*, Oxford University Press, Oxford, 2018.

³⁸ Cfr. C. CASONATO, *AI and Constitutionalism: The Challenges Ahead*, in *Reflections on Artificial Intelligence for Humanity*, a cura di B. BRAUNSCHWEIG, M. GHALLAB, Springer, Cham, https://dx.doi.org/10.1007/978-3-030-69128-8_9, 2021, 127–149; C. COLAPIETRO, *The Impact of the Information and Technology Revolution on the Constitutional Rights with Particular Attention to Personal Data Protection Issues*, in *The IT Revolution and its Impact on State, Constitutionalism and Public Law*, a cura di M. BELOV, Bloomsbury Publishing, London, 2021, 231–246.

³⁹ G. MALGIERI, *Automated Decision-making and Data Protection in Europe*, in *Research Handbook on Privacy and Data Protection Law*, a cura di G. GONZÁLEZ FUSTER, Edward Elgar Publishing, London, 2022, 433–448, 433.

⁴⁰ C.R. SUNSTEIN, *Algorithms, correcting biases*, in *Social Research: An International Quarterly*, 2, 2019, 499–511, 499.

⁴¹ B.A. HERRERA-TAPIAS, e al., *Algorithmic discrimination and explainable artificial intelligence in the judiciary: A case study of the Constitutional Court of Colombia*, cit.; I.N. COFONE, *Algorithmic discrimination is an information problem*, in *Hastings LJ*, 6, 2018, 1389–1444; F. ZUIDERVEEN BORGESUIS, *Discrimination, artificial intelligence, and algorithmic decision-making*, 2018 (reperibile in: <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>).

⁴² A. SIMONCINI, S. SUWEIS, *Il cambio di paradigma nell’intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Riv. fil. dir.*, 1, 2019, 87–106.

Alla luce di quanto indicato, il saggio intende approfondire in che modo i principi costituzionali relativi alla giustizia possano fungere da argine e guida all'implementazione dell'IA nel settore della giustizia. Si intende dimostrare che, tra tutti, il principio dell'umanità della decisione giudiziaria deriva dalla stessa Costituzione italiana e che non può mai essere limitato in nome dell'efficienza promessa dall'automazione della giustizia⁴³.

L'articolo seguirà la seguente argomentazione: dapprima si prenderà in esame la casistica e le analisi compiute fino alla comparsa sul mercato degli LLM (novembre 2022); successivamente verrà analizzato ciò che è accaduto negli ultimi tre anni dopo l'avvento degli LLM e, più in generale, della Generative AI (GenAI).

Si rileverà come, a decorrere dalla fine del 2022, sia intervenuto un cambiamento decisivo sia nella prassi sia nell'elaborazione concettuale di tali temi. In un lasso di tempo assai breve, la stessa immagine degli usi potenziali dell'intelligenza artificiale è risultata profondamente ridefinita. Non si è trattato soltanto di un progresso tecnologico, bensì dell'emersione di una nuova fase, nella quale la funzione giurisdizionale appare esposta a una tensione peculiare che, come si mostrerà, non è soltanto normativa e organizzativa, ma anche culturale. L'analisi si concluderà con la proposta di interpretare tale mutamento come l'avvio di una trasformazione strutturale della giustizia.

2. Gli anni dieci: l'impatto con i problemi del *Machine Learning*

Se si eccettuano alcuni studi sperimentali⁴⁴, la maggior parte della letteratura e degli studi pubblicati fino al 2022 che riguardano l'uso dell'IA nell'attività legale affermava che queste tecnologie non avrebbero potuto sostituire validamente l'attività umana nelle decisioni giudiziarie. Va detto che, fino a quel momento, vi erano pochi casi concreti che documentassero l'uso efficace dell'IA nei processi⁴⁵.

L'idea centrale che accompagnava gli impieghi dell'IA nell'ambito giudiziario era la "previsione (prediction) di esiti" tramite algoritmi di apprendimento automatico (Machine Learning)⁴⁶. Le situazioni tipo erano più o meno queste⁴⁷. Per determinare se una persona potesse essere un "buon" debitore, si procedeva alla sua profilazione, confrontandola automaticamente con altre persone in condizioni analoghe. Un'impostazione simile veniva adottata anche in altri ambiti, come nel parole system, relativo alla concessione della libertà su cauzione⁴⁸. Negli Stati Uniti, in particolare, il software COMPAS aveva acquisito

⁴³ Come tra l'altro indicato dall'art. 15 della legge n. 132/2025. Sul tema G. GALLONE, *Riserva di umanità, intelligenza artificiale e funzione giurisdizionale alla luce dell'IA Act. Considerazioni (e qualche proposta) attorno al processo amministrativo che verrà*, *Judicium*, 7 Novembre 2024 (reperibile in: <https://www.judicium.it/riserva-di-umanita-intelligenza-artificiale-e-funzione-giurisdizionale-alla-luce-dellia-act-considerazioni-e-qualche-proposta-attorno-al-processo-amministrativo-che-verra/>).

⁴⁴ Tra i lavori più interessanti sul punto si v. M. LIPPI, e al., *CLAUDETTE: An automated detector of potentially unfair clauses in online terms of service*, in *Artificial Intelligence and Law*, 2, 2019, 117–139.

⁴⁵ I.N. COFONE, *AI and judicial decision-making*, in *Artificial Intelligence and the Law in Canada*, a cura di F. MARTIN-BARITEAU, T. SCASSA, LexisNexis, Toronto, 2021, 1-20.

⁴⁶ B. ALARIE, e al., *How artificial intelligence will affect the practice of law*, in *Univ Toronto L J*, 1 (supplement), 2018, 106–124

⁴⁷ Per un'interessante indagine sugli usi allora censiti, si v. F. BORGESANO, e al., *Artificial intelligence and justice: a systematic literature review and future research perspectives on Justice 5.0*, in *European Journal of Innovation Management*, 11, 2025, 349–385.

⁴⁸ Per rispondere alla domanda sulla propensione di quella persona a "fuggire", si operava una profilazione con algoritmi in grado di confrontare le caratteristiche di un individuo con quelle di persone in condizioni analoghe durante il periodo di libertà vigilata. Questi algoritmi venivano utilizzati per prevedere i risultati a partire dai loro dati di addestramento. A differenza dell'IA basata su regole, che è costruita su istruzioni che indicano come elaborare i dati forniti, gli algoritmi di apprendimento automatico non sono costruiti con istruzioni specifiche. Il loro aspetto principale è quello di riconoscere "regolarità" all'interno di grandi quantità di dati. Il modello di IA è quindi capace di apprendere

ampia diffusione come strumento di valutazione del rischio di recidiva⁴⁹. Alcuni software di riconoscimento facciale erano utilizzati dai dipartimenti di polizia di tutto il mondo per identificare i sospetti prima dell'incriminazione e della condanna⁵⁰. Il governo canadese aveva preso in considerazione l'impiego di un'intelligenza artificiale (Tax Foresight) che determina la probabilità che una persona sia considerata un evasore fiscale, per influenzare le decisioni relative a contenziosi e verifiche fiscali. Analogamente, erano stati proposti diversi sistemi di intelligenza artificiale per i tribunali canadesi, con l'obiettivo di valutare le prove contraddittorie e prevedere l'esito dei processi⁵¹.

Al di là degli aspetti, pur molto importanti, relativi ai possibili profili discriminatori legati agli usi di queste tecnologie, di cui parleremo più avanti, l'effetto che questi esempi avevano determinato tra gli studiosi e nell'opinione pubblica era la paura che un giorno si sarebbe potuto sostituire il giudice con un algoritmo⁵².

La dottrina fondava tale impostazione sull'assunto dell'impossibilità di assimilare l'apprendimento automatico all'attività di giudizio. Il primo, infatti, opera mediante l'estrazione di correlazioni e ricorrenze dai dati; la seconda, per converso, si articola attraverso procedimenti di natura sillogistica e analogica. Se l'apprendimento automatico non è suscettibile di ricorrere, in senso proprio, al ragionamento per analogia, risolvendosi nell'individuazione di pattern ricorrenti, l'attività di giudizio postula invece la ricerca, la selezione e l'interpretazione di principi normativi comuni⁵³. A ciò si aggiunga che il tratto qualificante della funzione giurisdizionale è l'esercizio della discrezionalità interpretativa, che l'apprendimento automatico non può esprimere, in quanto inscindibilmente dipendente dal contesto di riferimento⁵⁴. Gli autori più avvertiti ravvisavano, sotto tale profilo, un'analogia tra l'impiego dell'apprendimento automatico — in tutte le sue declinazioni — ai fini decisionali nel processo e il suo impiego nell'esercizio dell'attività amministrativa⁵⁵.

Alcuni, a questo proposito, parlavano, forse in modo non del tutto esatto, di “giustizia predittiva”, con espressione intesa a designare l'impiego di strumenti algoritmici idonei a formulare previsioni sull'esito della decisione, ma non per questo capaci di surrogare il proprium della funzione giurisdizionale⁵⁶.

le proprie regole decisionali per prevedere o stimare qualcosa. K.D. ASHLEY, S. BRÜNINGHAUS, *Automatically classifying case texts and predicting outcomes*, cit..

⁴⁹ V. *infra* in questo paragrafo.

⁵⁰ G. MOBILIO, *Tecnologie di riconoscimento facciale. Rischi per i diritti fondamentali e sfide regolative*, Editoriale Scientifica, Napoli, 2021.

⁵¹ C. INTAHCHOMPHOO, e al., *References to Artificial Intelligence in Canada's Court Cases*, in *Legal Information Management*, 1, 2020, 39–46.

⁵² T. SOURDIN, *Justice and technological Innovation*, in *Journal of Judicial Administration*, 2, 2015, 96–105.

⁵³ C.R. SUNSTEIN, *Of artificial intelligence and legal reasoning*, in *U. Chi. L. Sch. Roundtable*, 1, 2001, 29–36.

⁵⁴ R.M. RE, A. SOLOW-NIEDERMAN, *Developing artificially intelligent justice*, in *Stan. Tech. L. Rev.*, 2, 2019, 242–289; C. COGLIANESE, D. LEHR, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, in *Geo. LJ*, 2016, 1147–1223.

⁵⁵ Il divieto di impiegare l'IA nell'attività amministrativa fu oggetto di una famosa vicenda giudiziaria italiana legata a un algoritmo utilizzato dal Ministero dell'istruzione per distribuire alcuni insegnanti in base a una graduatoria. Siamo parlando del famoso caso relativo all'algoritmo della “Buona Scuola”. Su di esso esiste una letteratura sconfinata. *Infra multis* si v. no A. SIMONCINI, *Amministrazione digitale algoritmica. Il quadro costituzionale*, in *Il diritto dell'amministrazione pubblica digitale*, a cura di R. CAVALLO PERIN, D.-U. GALLETTA, Giappichelli, Torino, 2020, 1–42;

⁵⁶ Ampiamente sul tema: S. LEBRETON-DERRIEN, *La justice prédictive. Introduction à une justice "simpement" virtuelle*, in *Arch. phil. droit*, 2018, 1–20; F. GALLI, G. SARTOR, *AI Approaches to Predictive Justice: A Critical Assessment*, in *HUMANITIES AND RIGHTS GLOBAL NETWORK JOURNAL*, 2, 2023, 165–217; C. CASTELLI, D. PIANA, *Giustizia predittiva. La qualità della giustizia in due tempi*, in *Quest. giust.*, 4, 2018, 153–165; A. DE LA OLIVA SANTOS, “Giustizia predittiva”, *interpretazione matematica delle norme, sentenze robotiche e la vecchia storia del “Justizklavier”*, in *Riv. trim. dir. proc. civ.*, 2019, 883–895; S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale “predittiva”*, in *Cassazione penale*, 4, 2019,

Da tali vicende emergeva anche un altro importante dato: affidarsi all'IA per prendere decisioni non significava che quest'ultima potesse emulare i modi di ragionare dei giuristi⁵⁷, e men che meno le attività processuali. L'idea centrale alla base di questa distinzione è che l'attività giudiziaria è un processo esclusivamente umano, con caratteristiche rituali, che varia notevolmente a seconda del contesto di riferimento⁵⁸. La differenza tra macchine ed esseri umani è notevole⁵⁹. È diverso il modo in cui elaborano le informazioni e giungono a conclusioni. I giudici esseri umani prendono decisioni basandosi su conoscenze giuridiche e culturali, sull'esperienza, sull'interazione con le competenze di altri (avvocati, periti nominati dal tribunale, testimoni, ecc.), sul contesto e, talvolta, sul buon senso⁶⁰. Nessuna macchina può emulare tale modo di fare né essere autonoma in queste attività⁶¹.

Alla IA – si diceva – manca la capacità di innovare o di modificare le leggi adattandole all'evoluzione delle norme sociali o culturali⁶². E se anche lo sapesse fare, non capirebbe perché lo sta facendo.

Questi studi, oltretutto, riconoscevano sempre che non sarebbe auspicabile il controllo delle attività giudiziarie da parte di una macchina, giacché ne sarebbe derivato sia un cambiamento dei meccanismi propri dello *ius dicere* sia uno stravolgimento dei valori giudiziari alla base del sistema giuridico⁶³.

Un altro tema di discussione è la differenza tra le disposizioni giuridiche e il linguaggio informatico (i cd. codici)⁶⁴. Lo stesso testo di una disposizione giuridica (prima di esaminarne l'interpretazione e l'applicazione) non è facilmente traducibile in codice binario né tantomeno assimilabile all'inferenza statistica dell'apprendimento automatico. Sia i primi studiosi del rapporto tra diritto e informatica sia quelli del rapporto tra diritto e IA si erano fermati di fronte alla banale traduzione dei principi giuridici, come quello del “buon padre di famiglia” o dell’“ordine pubblico”, in codice binario⁶⁵.

Le disposizioni giuridiche vengono quasi sempre redatte richiamando principi che hanno uno spettro di significati molto ampio⁶⁶, il cui valore semantico è funzionale a garantirne l'applicazione anche oltre lo specifico caso e, in caso di bilanciamento con altri

1748–1765; G. SARTOR, A. SANTOSUOSSO, *La giustizia predittiva: una visione realistica*, in *Giur. it.*, 7, 2022, 1760–1782.

⁵⁷ H. SURDEN, *Machine Learning and Law*, in *Wash. L. Rev.*, 1, 2014, 87–115.

⁵⁸ Si v. no a tal proposito quanto scrivono: A. GARAPON, J. LASSÈGUE, *Justice Digitale. Révolution graphique et rupture anthropologique*, Presses Universitaires de France/Umensis, Paris, trad. it. *La giustizia digitale. Determinismo tecnologico e libertà*, il Mulino, Bologna, il Mulino; J. MORISON, A. HARKENS, *Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making*, cit..

⁵⁹ C. BONA, *Sentenze imperfette: gli errori cognitivi nei giudizi civili*, il Mulino, Bologna, 2010, 202 ss.

⁶⁰ Sul punto v. quanto scrive in generale R.A. POSNER, *Reflections on judging*, Harvard University Press, 2013, 54 ss. e invece quanto argomenta nello specifico T. SOURDIN, *Robo Justice: Constitutional Issues with Judge AI*, in *Ind. J. Global Legal Stud.*, 2, 2023, 293–324.

⁶¹ D.A. LARSON, *Artificial Intelligence: Robots, avatars, and the demise of the human mediator*, in *Ohio St. J. on Disp. Resol.*, 1, 2010, 105–164

⁶² F. PASQUALE, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, in *Geo. Wash. L. Rev.*, 1, 2019, 1–55.

⁶³ I.N. COFONE, *AI and judicial decision-making*, cit., 5.

⁶⁴ J. OSTER, *Code is code and law is law—the law of digitalization and the digitalization of law*, in *International Journal of Law and Information Technology*, 2, 2021, 101–117.

⁶⁵ Come ricordano A. SANTOSUOSSO, G. SARTOR, *Decidere con l'IA. Intelligenze artificiali e naturali nel diritto*, il Mulino, Bologna, 2024, p. 22 ss. Sul punto sia consentito rimandare anche a E. LONGO, *Giustizia digitale e Costituzione: riflessioni sulla trasformazione tecnica della funzione giurisdizionale*, cit.

⁶⁶ Su questo tema v. J. WALDRON, *Vagueness in Law and Language: Some Philosophical Issues*, in *California Law Review*, 3, 1994, 509–540; L. LOMBARDI VALLAURI, *Norme vaghe e teoria generale del diritto*, in *Jus*, 1, 1999, 25–32; H. ASGEIRSSON, *The nature and value of vagueness in the law*, Hart, Oxford, 2020.

principi⁶⁷, a garantirne un'applicazione proporzionale. Lo sanno bene i giudici. L'interpretazione e l'applicazione delle fonti del diritto comportano decisioni discrezionali, nelle quali si comprende che il ragionamento necessario non è neanche lontanamente simile a un'operazione matematico-statistica⁶⁸. Un giudice deve prima di tutto ricostruire il caso, applicare ad esso una norma, risolvere una controversia e contribuire a chiarire il significato attuale della norma giuridica oggetto del suo giudizio. Lo stesso si può dire degli avvocati e di chiunque opera professionalmente nell'ambito del diritto⁶⁹.

2.1. L'uso "ausiliario" dell'IA e i problemi derivanti dallo "human in the loop"

L'integrazione dell'IA in un'attività guidata dall'uomo è spesso descritta con l'espressione "human in the loop" (HITL)⁷⁰. Tale formula indica che il ciclo decisionale in cui entra in scena l'IA, chiamato anche "automated decision-making"⁷¹, prevede forme di supervisione o di intervento umano⁷². In senso ampio, si parla di HITL quando l'essere umano può influenzare il funzionamento o l'esito del sistema (ad esempio attraverso correzioni, *feedback* o validazioni) oppure, una volta prodotto l'output, scegliere se seguirlo o ignorarlo⁷³.

Talvolta il ruolo dell'essere umano è prevalentemente procedurale, ad esempio nel decidere quali casi meritino maggiori risorse istituzionali o quale canale istruttorio attivare. In altri casi, invece, il ruolo umano è più determinante e include il potere di incidere sugli esiti: sia in termini di effetti concreti sul destinatario (ad esempio concedere o negare un beneficio), sia in termini di giustificazione della decisione, o entrambi. Le specificità del ruolo umano possono variare, ma l'elemento decisivo è che l'essere umano conservi una discrezionalità significativa in determinate fasi o in determinati casi⁷⁴.

⁶⁷ Come rilevano F. PASQUALE, G. CASHWELL, *Four futures of legal automation*, in *UCLA L. Rev. Discourse*, 1, 2015, 26–48, le attività di maggiore complessità relative alla applicazione della legge non potevano essere automatizzate a meno di non limitare lo stesso significato e l'efficacia delle norme giuridiche.

⁶⁸ Anche gli autori più avveduti erano arrivati alla conclusione della assoluta complessità della legge e della sua "irriducibilità" al codice. Cfr. J. OSTER, *Code is code and law is law—the law of digitalization and the digitalization of law*, cit..

⁶⁹ B. ALARIE, A. NIBLETT, A.H. YOON, *How artificial intelligence will affect the practice of law*, cit..

⁷⁰ La letteratura sul tema è vasta. *Infra multis* si v. R. CROTOF, e al., *Humans in the Loop*, in *Vand. L. Rev.*, 2, 2023, 429–510; E. MOSQUEIRA-REY, e al., *Human-in-the-loop machine learning: a state of the art*, in *Artificial Intelligence Review*, 2022, 3005–3054; L.F. CRANOR, *A framework for reasoning about the human in the loop*, in E. CHURCHILL, R. DHAMIJA, 2008 (reperibile in: <https://dl.acm.org/doi/proceedings/10.5555/1387649>).

⁷¹ Si v. K. BRENNAN-MARQUEZ, e al., *Strange Loops: Apparent Versus Actual Human Involvement in Automated Decision Making*, in *Berkeley Technology Law Journal*, 3, 2019, 745–772.

⁷² F. PASQUALE, *Inalienable Due Process in an Age of AI: Limiting the Contractual Creep toward Automated Adjudication*, in *Constitutional Challenges in the Algorithmic Society*, a cura di H.-W. MICKLITZ, et al., Cambridge University Press, Cambridge, <https://www.cambridge.org/core/books/constitutional-challenges-in-the-algorithmic-society/constitutional-law-in-the-algorithmic-society/969E0889109C8092AD0AB57019E507E3>, 2022, 42–56; F. PALMIOTTO, *When Is a Decision Automated? A Taxonomy for a Fundamental Rights Analysis*, in *German Law Journal*, 2, 2024, 210–236.

⁷³ In realtà, la letteratura distingue diverse configurazioni di "presenza" umana: nello HITL l'umano interviene durante il processo (ad esempio validando passaggi, correggendo l'output, fornendo etichette, dando *feedback* che modifica il comportamento del sistema); nello *human-on-the-loop* (HOTL) l'umano non opera necessariamente "dentro" il processo in modo continuativo, ma svolge un ruolo di monitoraggio e controllo, con la possibilità di intervenire *ex post* o in caso di anomalia (supervisione, audit, *override*); nello *human-out-of-the-loop* (HOOTL) il sistema opera in modo sostanzialmente autonomo, senza un intervento umano significativo né in fase di esecuzione né (talvolta) in fase di decisione finale; l'umano, se presente, resta ai margini o svolge funzioni solo formali.

⁷⁴ Il software COMPAS, ad esempio, produceva una previsione di rischio, ma, in ultima analisi, spettava al giudice decidere se seguirla. T. BRENNAN, W. DIETERICH, *Correctional offender management profiles for alternative sanctions (COMPAS)*, in *Handbook of recidivism risk/needs assessment tools*, a cura di J.P. SINGH, et al., Wiley, New York, 2018, 49–75

Pensare all'essere umano "nel loop" significa dunque riformulare l'adozione della tecnologia, allontanandosi dall'idea della sostituzione e orientandosi verso la progettazione e l'implementazione di strumenti che assistano i decisori umani (nel nostro caso, i giudici). In questa prospettiva, la sfida diventa definire la portata concreta del ruolo umano nella relazione con l'IA: se e in che misura il giudice debba essere *in*, *on* o, di fatto, *out* "of the loop", e, soprattutto, fino a che punto sia opportuno affidarsi alle raccomandazioni del sistema.

Astrattamente l'uso "assistivo" dell'IA nel processo può effettivamente aumentare l'efficienza e l'uniformità delle operazioni. L'IA può essere più rapida, meno costosa e, per alcuni compiti, anche più accurata degli esseri umani. Tuttavia, affinché tale integrazione sia efficace, occorre un livello adeguato di "deferenza" nei confronti della macchina, evitando due rischi: la sottovalutazione sistematica dell'output (sfiducia) o la sua accettazione acritica. Vediamo cosa accade in tali casi.

Se un giudice non si fida dei risultati di un sistema di IA, quel sistema diventa sostanzialmente inutile: le previsioni algoritmiche hanno scarso valore se vengono sistematicamente ignorate. La tendenza a perdere fiducia in un sistema dopo aver constatato un errore è spesso descritta come "avversione algoritmica"⁷⁵.

All'opposto, la tendenza a riporre eccessiva fiducia nei risultati del sistema può determinare un effetto ancora più problematico. Se un giudice si affida alle conclusioni dell'algoritmo senza confrontarle con altre prove, con la propria esperienza e con il contesto culturale e giuridico, l'integrazione dell'IA può finire per compromettere il processo decisionale⁷⁶. Balkin spiega tale fenomeno con il noto "effetto di sostituzione"⁷⁷.

⁷⁵ Nella letteratura straniera il tema dell'algorithmic aversion, anche nella forma della "resistance", è molto presente. Si v. no *infra multis*: B. DIETVORST, e al., *Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err*, in *Journal of Experimental Psychology: General*, 1, 2015, 114–126; J. VELKOVA, A. KAUN, *Algorithmic resistance: Media practices and the politics of repair*, in *Information, Communication & Society*, 4, 2021, 523–540; B.J. DIETVORST, e al., *Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them*, in *Management science*, 3, 2018, 1155–1170; E. JUSSUPOW, e al., *Why are we averse towards Algorithms? A comprehensive literature Review on Algorithm aversion*, in An Online AIS Conference, 2020 (reperibile in: https://aisel.aisnet.org/ecis2020_rp/168); C. COGLIANESE, A. LAI, *Algorithm vs. algorithm*, in *Duke LJ*, 6, 2021, 1281–1340; H. MAHMUD, e al., *What influences algorithmic decision-making? A systematic literature review on algorithm aversion*, in *Technological forecasting and social change*, 2022, 121390; B.M. CHEN, e al., *Having your day in robot court*, in *Harv. JL & Tech.*, 1, 2022, 127–169.

⁷⁶ Questa tendenza è nota come *automation bias* (pregiudizio dell'automazione) e riguarda il modo in cui i decisori umani trattano gli output dei sistemi automatizzati (a differenza dei pregiudizi algoritmici, che riguardano il sistema in sé). Sul tema la letteratura è sterminata. Si v. no in particolare: R. PARASURAMAN, D.H. MANZEY, *Complacency and Bias in Human Use of Automation: An Attentional Integration*, in *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 3, 2010, 381–410; K. GODDARD, e al., *Automation bias: a systematic review of frequency, effect mediators, and mitigators*, in *Journal of the American Medical Informatics Association*, 1, 2012, 121–127; S. ALON BARKAT, M. BUSUIOC, *Human-AI Interactions in Public Sector Decision-Making: 'Automation Bias' and 'Selective Adherence' to Algorithmic Advice*, in *Journal of Public Administration Research and Theory*, 1, 2022, 153–169. In una certa misura, per realizzare un'efficace integrazione uomo-macchina, è necessario che la fiducia sia correttamente calibrata. Sebbene l'avversione algoritmica possa manifestarsi quando gli errori diventano evidenti, i decisori mostrano spesso una marcata propensione al pregiudizio nei confronti dell'automazione, talvolta con esiti gravi. Vedi: R. CROTOF, *"Cyborg Justice" and the Risk of Technological-Legal Lock-in*, in *Colum. L. Rev. Online*, 1, 2019, 233–251; M.C. ELISH, *Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction*, in *Engaging Science, Technology, and Society*, 2019, 40–60.

⁷⁷ Attraverso l'interazione con sistemi di IA, le persone sono disposte a delegare loro compiti che non desiderano più svolgere direttamente, sostituendoli, in determinati contesti e per determinati scopi, ad esseri umani o animali. Tale delega, tuttavia, è sempre parziale, perché riguarda un compito e non il ruolo complessivo della persona. Gli esseri umani non vengono mai completamente sostituiti dall'IA, ma possono usare l'IA per svolgere alcuni compiti che prima svolgevano loro stessi (o altri). I rischi specifici dell'uso dell'IA, sostiene Balkin, derivano soprattutto da una sostituzione involontaria degli esseri umani con le macchine in compiti specifici. V. J.M. BALKIN, *The Path of Robotics Law*, in *California Law Review Circuit*, 1, 2015, 45–60.

Alcuni usi meno recenti di algoritmi di Machine Learning nel processo hanno messo in evidenza un altro problema molto importante quando si delega parte del processo decisionale giudiziario a una macchina⁷⁸. Questi esempi, almeno per come sono stati elaborati dalla dottrina, non descrivono fenomeni di sostituzione degli esseri umani da parte di un'IA. Anche quando l'IA subentra nell'esecuzione di un singolo compito prima svolto da una persona, non sostituisce il giudice come figura istituzionale responsabile. Piuttosto, al giudice umano viene richiesto di svolgere un ruolo diverso: il baricentro si sposta da chi esegue direttamente il compito a chi progetta, seleziona, applica e supervisiona la tecnologia. In ogni caso, resta sempre un essere umano che, in qualche modo, assume decisioni o le rende operative, anche quando tale ruolo è debole o meramente formale⁷⁹. Data la natura dell'attività che l'ordinamento disciplina, un essere umano resta comunque coinvolto, ma la questione decisiva diventa la modifica delle relazioni sociali – in questo caso tra decisori e cittadini – quando vengono mediate da tali tecnologie⁸⁰.

2.2. Due rischi chiave nell'uso assistivo dell'IA nel processo

Nella misura in cui i giudici si avvalgono dell'IA in forma assistiva (nelle varie configurazioni di HITL, HOTL e simili già descritte), emergono due rischi chiave per l'amministrazione della giustizia: il *pregiudizio algoritmico* (*algorithmic bias*) e l'*opacità* (*opacity*).

Il pregiudizio algoritmico affonda le sue radici nella condizione umana stessa: tutti gli individui, anche i più animati da buone intenzioni, sono soggetti a pregiudizi, inclusi quelli impliciti, che possono manifestarsi in qualsiasi processo decisionale. Uno dei motivi per cui gli algoritmi vengono impiegati è proprio la loro presunta neutralità, che li renderebbe capaci di superare tali distorsioni. Tuttavia, esiste un'ampia letteratura che confuta l'idea di imparzialità algoritmica, corroborata da numerosi esempi e casi documentati in cui algoritmi decisionali hanno prodotto risultati distorti e discriminatori⁸¹. È fondamentale

⁷⁸ In Estonia sono stati sperimentati strumenti algoritmici per la gestione di controversie di modesta entità; analogamente, negli Stati Uniti, “DoNotPay”, la società tech che offre consulenza legale promuovendo il suo chatbot AI offre servizi di consulenza automatizzata di base, come il supporto per contestare multe o l'analisi di contratti standard. Un altro esempio ben noto è quello dell'azienda “AirHelp”, che, utilizzando sistemi di intelligenza artificiale, assiste i passeggeri nelle loro richieste di risarcimento alle compagnie aeree. La stessa azienda spiega che i suoi “avvocati digitali”, in grado di controllare i documenti di viaggio, effettuare valutazioni legali e analizzare la giurisdizione per una richiesta di risarcimento in pochi secondi, vengono utilizzati nel 100% delle richieste in arrivo nella fase legale, e quelle che vengono approvate in questa fase vengono passate a un secondo *bot* che inizia ad analizzare la richiesta. I dati dell'azienda riportano che questo algoritmo è già stato utilizzato con successo in oltre 150.000 casi e l'esito è stato positivo nel 96% dei casi, il 5% in più rispetto al tasso medio di successo umano. Per questi casi, si v. M.G. CIVININI, *The Perks of Using AI Tools in the Justice Domain*, cit., 30; T. SOURDIN, e al., *Digital technology and justice: the use of justice apps*, Routledge, New York, NY, 2021, 1 ss.

⁷⁹ T. ENARSSON, e al., *Approaching the human in the loop – legal perspectives on hybrid human/algorithmic decision-making in three contexts*, in *Information & Communications Technology Law*, 1, 2022, 123–153.

⁸⁰ La letteratura sul rapporto tra pubblica amministrazione e cittadini sta già affrontando questo problema: S. RANCHORDÁS, *Empathy in the digital administrative state*, in *Duke LJ*, 6, 2021, 1341–1390; U.B.U. ROEHL, M.B. HANSEN, *Automated, administrative decision-making and good governance: Synergies, trade-offs, and limits*, in *Public Administration Review*, 6, 2024, 1184–1199.

⁸¹ Prima ancora che emergesse il caso COMPAS, già alcuni Autori si erano posti il problema. V. K. CRAWFORD, J. SCHULTZ, *Big data and due process: Toward a framework to redress predictive privacy harms*, in *BCL Rev.*, 1, 2014, 93–128. In generale sul tema si v. F. PASQUALE, *Inalienable Due Process in an Age of AI: Limiting the Contractual Creep toward Automated Adjudication*, cit., 46; N. KORDZADEH, M. GHASEMAGHAEI, *Algorithmic bias: review, synthesis, and future research directions*, in *European Journal of Information Systems*, 3, 2022, 388–409.

comprendere le implicazioni del pregiudizio algoritmico diventa essenziale: può infatti condurre a esiti non solo inaccurati, ma anche profondamente iniqui⁸².

Per il vero, l'idea stessa di decisioni algoritmiche eque e imparziali è, in sé, un errore. Ciò non significa che gli algoritmi non siano utili per assistere gli esseri umani nel processo decisionale. Lo sono. Ma è necessario definire un quadro normativo che prenda in considerazione tali pregiudizi, distinguendoli da quelli che affliggono gli esseri umani. Se non controllati, questi pregiudizi, riguardanti o l'algoritmo stesso⁸³, o i dati⁸⁴, o i pregiudizi sociali amplificati⁸⁵, possono indurre i giudici ad avallare, inavvertitamente, risultati inaccurati o discriminatori nelle loro sentenze.

Come già detto in precedenza, il momento in cui gran parte della letteratura giuridica si è accorta di questi problemi è stato con il già citato caso *Loomis*⁸⁶ e con l'indagine

⁸² A.D. SELBST, e al., *Fairness and abstraction in sociotechnical systems*, in Atlanta, GA, 2019 (reperibile in: <https://dl.acm.org/doi/pdf/10.1145/3287560.3287598>).

⁸³ Il primo tipo di pregiudizio sorge durante il processo di costruzione del modello algoritmico (bias di costruzione). Si tratta di un pregiudizio nel modo in cui un algoritmo elabora le informazioni. I bias nel processo algoritmico spesso esistono perché i pregiudizi umani vengono “tradotti” nel sistema. Anche se nessun essere umano sceglie direttamente l'esito, c'è sempre un coinvolgimento umano nel modo in cui quell'esito viene raggiunto: gli esseri umani definiscono il problema e decidono cosa l'algoritmo debba prevedere prima ancora che qualunque dato venga elaborato. Una volta presa questa decisione, l'intervento umano continua nella raccolta dei dati per addestrare l'algoritmo e nella selezione delle variabili che l'algoritmo deve considerare (le “feature”). In qualche misura, quindi, ci sono sempre decisori umani che influenzano il processo. In letteratura si v. I.N. COFONE, *Algorithmic discrimination is an information problem*, cit.; C.R. SUNSTEIN, *Algorithms, correcting biases*, cit.

⁸⁴ Il secondo tipo di pregiudizio ha origine nel campione di dati utilizzato per addestrare l'algoritmo (bias del campione o del dataset). Il potere predittivo di un algoritmo è “buono” solo quanto lo sono i dati che riceve. Se l'algoritmo “scava” in una porzione del dataset che, per qualsiasi ragione, non è rappresentativa della popolazione, produrrà risultati non rappresentativi (cioè decisioni individuali inaccurate e potenzialmente discriminatorie). Ciò accade perché i singoli record possono avere problemi di qualità, ad esempio perché i dati sono parziali o errati. Può accadere che anche l'intero dataset possa avere problemi: può essere di qualità peggiore, a tassi più elevati, per un determinato gruppo protetto rispetto ad altri, oppure può non rappresentare la popolazione generale. K. MAVROGIORGOS, e al., *Bias in Machine Learning: A Literature Review*, in *Applied Sciences*, 19, 2024, 8860; N. ZHOU, e al., *Bias, Fairness and Accountability with Artificial Intelligence and Machine Learning Algorithms*, in *International Statistical Review*, 2, 2022, 1–13; N. KORDZADEH, M. GHASEMAGHAEI, *Algorithmic bias: review, synthesis, and future research directions*, cit.

⁸⁵ Il terzo tipo di pregiudizio riguarda i pregiudizi sociali, che l'algoritmo cattura e amplifica. I dati di addestramento di un sistema di machine learning possono incorporare discriminazioni sistemiche pregresse. Di conseguenza, un'IA può produrre un impatto disparato o una discriminazione indiretta anche quando è addestrata correttamente con dati rappresentativi. La differenza rispetto al caso del “campione distorto” è che qui i dati sono rappresentativi della popolazione, ma producono comunque un esito a impatto disparato perché rispecchiano disuguaglianze sociali già radicate. S.-H. HONG, *AI and bias*, in *Handbook on Public Policy and Artificial Intelligence*, a cura di R. PAUL, et al., Edward Elgar Publishing, London, 2024, 109–122;

⁸⁶ *State v. Loomis*. 881 N.W.2d 749 (Wis. 2016). Nel luglio 2016, la Corte Suprema del Wisconsin ha confermato l'utilizzo dell'algoritmo “Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)” di Northpointe Inc. nel processo decisionale giudiziario concernente l'imputato Eric Loomis. Loomis era stato valutato da COMPAS come individuo ad alto rischio e di conseguenza condannato a otto anni di carcere, una sentenza che egli ha contestato come violazione dei suoi diritti processuali. Sebbene i risultati della sua “valutazione del rischio” gli fossero stati comunicati, i calcoli che avevano trasformato i dati sottostanti in tale punteggio non erano stati rivelati e il suo avvocato aveva chiesto al Tribunale di imporre a Northpointe di rivelarli, senza però ottenere giustizia nel caso specifico.

Un esame approfondito del caso Loomis e di altri casi giurisprudenziali ha contribuito a dimostrare i rischi associati alla algoritmizzazione su larga scala della giustizia, insieme ai limiti della “scatola nera tecnica” nella protezione dei diritti fondamentali quali il giusto processo, la trasparenza e l'uguaglianza. Solo dopo un'approfondita inchiesta giornalistica condotta, tra gli altri, da *Pro Publica*, il software è stato pubblicamente accusato di essere discriminatorio, in particolare nei confronti degli individui afroamericani. L'indagine ha rivelato che gli imputati di colore avevano il doppio delle probabilità rispetto agli imputati bianchi di essere classificati in modo errato come ad alto rischio da COMPAS. Inoltre, gli imputati bianchi erano più propensi ad essere erroneamente classificati come a basso rischio rispetto agli imputati di colore.

Ciononostante, la Corte Suprema del Wisconsin non ha ritenuto legittimo riesaminare il caso alla luce dei rilievi critici sollevati, né ha considerato che tali elementi critici fossero almeno sufficienti a riformare la decisione presa in prima

sull'impiego del software COMPAS⁸⁷ per la valutazione del rischio nei procedimenti penali negli Stati Uniti, alla quale è seguita una vastissima letteratura tecnica sui *bias* algoritmici⁸⁸.

La seconda critica, comune all'uso degli algoritmi nella giustizia, riguarda l'"opacità" delle decisioni. Queste ultime sono difficili o impossibili da comprendere, sia per la loro complessità sia per le protezioni della proprietà intellettuale che presidia il cd. "codice sorgente"⁸⁹.

L'opacità può essere vista sia in termini positivi, come esigenza di spiegabilità e trasparenza⁹⁰, sia in termini negativi, come mancanza di conoscenza o di comprensione da parte degli individui su come un algoritmo sia giunto a un risultato (la decisione) a partire dai suoi input (i dati che gli vengono forniti). Comprendere l'opacità algoritmica è importante sia per chiarire la portata degli interessi privati nel mantenere segreti i sistemi di IA impiegati nell'amministrazione della giustizia, sia per comprendere la natura stessa della tecnologia alla base dell'IA.

La mancanza di trasparenza è un grande problema quando si usa l'IA nel processo decisionale giudiziario, perché i sistemi di IA – come abbiamo visto prima in relazione ai

istanza, ignorando di fatto le accuse di effetti discriminatori, il tutto partendo dal presupposto che non si può ravvisare alcuna violazione del giusto processo se il sistema è stato applicato correttamente e se altri fattori corroborano la valutazione automatica. Per una sintesi del caso si veda J. LARSON, e al., *How we analyzed the COMPAS recidivism algorithm*, ProPublica, 23/5/2023 (reperibile in: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>); D.L. KEHL, S.A. KESSLER, *Algorithms in the criminal justice system: Assessing the use of risk assessments in sentencing*, 2017 (reperibile in: <https://dash.harvard.edu/handle/1/33746041>); A.Z. HUQ, *Racial equity in algorithmic criminal justice*, in *Duke LJ*, 6, 2018, 1043–1134; T. BRENNAN, W. DIETERICH, *Correctional offender management profiles for alternative sanctions (COMPAS)*, cit., ; G. SARTOR, F. LAGIOIA, *Il sistema COMPAS: algoritmi, previsioni, iniquità*, in *XXVI Lezioni di Diritto dell'Intelligenza Artificiale*, a cura di U. RUFFOLO, Giappichelli, Torino, 2021, 226–244.

⁸⁷ Su questo software ci siamo già soffermati. Tra l'altro, l'uso del software COMPAS non è un caso isolato. Altri strumenti di valutazione del rischio utilizzati nella giustizia penale mostrano risultati simili, con distorsioni analoghe. Ad esempio, uno strumento utilizzato a livello federale in Canada per prendere decisioni in materia di libertà vigilata (*Post Conviction Risk Assessment*) è risultato assegnare ai criminali di colore punteggi di valutazione del rischio post-condanna più elevati rispetto ai criminali bianchi. L'utilizzo dei precedenti penali come fattore predittivo riflette i pregiudizi sociali (il terzo tipo di pregiudizio algoritmico sopra menzionato), poiché i precedenti penali riflettono il rapporto tra razza e arresto, dove gli individui di colore sono più soggetti ad essere arrestati rispetto agli individui bianchi a parità di livello di attività criminale. J.L. SKEEM, C.T. LOWENKAMP, *Risk, race, and recidivism: Predictive bias and disparate impact*, in *Criminology*, 4, 2016, 680–712.

⁸⁸ Si v. *infra multis*, J. DRESSSEL, H. FARID, *The accuracy, fairness, and limits of predicting recidivism*, in *Science advances*, 1, 2018, eaao5580; A.L. WASHINGTON, *How to argue with an algorithm: Lessons from the COMPAS-ProPublica debate*, in *Colo. Tech. LJ*, 1, 2018, 131–160; A.Z. HUQ, *Racial equity in algorithmic criminal justice*, cit.; T. BRENNAN, W. DIETERICH, *Correctional offender management profiles for alternative sanctions (COMPAS)*, cit., ; D.L. KEHL, S.A. KESSLER, *Algorithms in the criminal justice system: Assessing the use of risk assessments in sentencing*; J. LARSON, S. MATTU, L. KIRCHNER, J. ANGWIN, *How we analyzed the COMPAS recidivism algorithm*, cit.; T. BRENNAN, e al., *Evaluating the predictive validity of the COMPAS risk and needs assessment system*, in *Criminal Justice and behavior*, 1, 2009, 21–40; A. CHANDER, *The racist algorithm*, in *Mich. L. Rev.*, 6, 2016, 1023–1046.

⁸⁹ Esiste una letteratura ancora più sconfinata su questo tema. Oltre agli articoli su aspetti puntuali dell'opacità, si v. no: B. VAASSEN, *AI, Opacity, and Personal Autonomy*, in *Philosophy & Technology*, 4, 2022, 1–20; F. PASQUALE, *Inalienable Due Process in an Age of AI: Limiting the Contractual Creep toward Automated Adjudication*, cit., 42 ss.; F. PALMIOTTO, *The Black Box on Trial: The Impact of Algorithmic Opacity on Fair Trial Rights in Criminal Proceedings*, in *Algorithmic Governance and Governance of Algorithms*, a cura di Springer, 2021, 49–70; J. BURRELL, *How the machine 'thinks': Understanding opacity in machine learning algorithms*, in *Big Data & Society*, 1, 2016, 2053951715622512; Y. VAN DEN EEDE, *In between us: On the transparency and opacity of technological mediation*, in *Foundations of Science*, 2-3, 2011, 139–159; R. WEXLER, *Life, liberty, and trade secrets: Intellectual property in the criminal justice system*, in *Stan. L. Rev.*, 5, 2018, 1343–1429; T. DANCY, M. ZALNIERIUTE, *AI and Transparency in Judicial Decision Making*, in *Oxford Journal of Legal Studies*, 2025, 1–34.

⁹⁰ L'opacità interferisce con il diritto delle persone soggette alle decisioni delle macchine a una spiegazione: si v. S. WACHTER, e al., *Why a right to explanation of automated decision-making does not exist in the general data protection regulation*, in *International Data Privacy Law*, 2, 2017, 76–99; M.E. KAMINSKI, *The Right to Explanation, Explained*, in *Berkeley Tech. L.J.*, 1, 2019, 189–218.

diversi tipi di bias – possono essere distorti in modi diversi rispetto agli esseri umani, di cui possono sostituire le funzioni⁹¹. Va ricordato che l’opacità complica l’applicazione delle dottrine che tutelano i diritti civili delle persone, come quella della discriminazione indiretta, poiché è difficile correggere un processo decisionale a cui non si può accedere né comprenderlo⁹². Perciò, la divulgazione dei processi algoritmici utilizzati nel processo decisionale giudiziario è importante per tutelare i diritti civili delle persone, promuovere la responsabilità e garantire l’equità procedurale⁹³.

Al di là del fatto che l’opacità deriva anzitutto dalla qualità dei dati, perché la mancanza di trasparenza sulla provenienza e sulla cura dei set di addestramento oscura la “logica dell’input” prima ancora che venga eseguita una singola riga di codice, la letteratura distingue l’opacità nei sistemi di IA in base a tre paradigmi ricorrenti⁹⁴: (i) l’opacità istituzionale (intenzionale)⁹⁵, (ii) l’opacità di competenza o comunicazione⁹⁶ e (iii) l’opacità tecnico-epistemica⁹⁷.

Ogni tipo di opacità richiede una risposta specifica sia dal sistema giudiziario sia dai giudici. L’opacità istituzionale (intenzionale) può essere ridotta attraverso obblighi di divulgazione mirati, ad esempio richiedendo l’accesso alla documentazione relativa a caratteristiche, convalida, tassi di errore e processi di governance, oppure fornendo accesso riservato a tribunali, autorità di regolamentazione o revisori indipendenti⁹⁸. Tali misure possono aiutare a valutare se le norme antidiscriminatorie e i requisiti di giusto processo siano rispettati, senza necessariamente richiedere la divulgazione completa al pubblico dei dettagli proprietari⁹⁹.

L’opacità dovuta alla mancanza di competenze o conoscenze specialistiche può essere affrontata al meglio attraverso meccanismi che traducono le informazioni tecniche in forme legalmente utilizzabili¹⁰⁰.

⁹¹ L’opacità rende più difficile individuare e ridurre tali distorsioni algoritmiche, come evidenzia M.C. BUITEN, *Towards Intelligent Regulation of Artificial Intelligence*, in *European Journal of Risk Regulation*, 1, 2019, 41–59.

⁹² C. O’NEIL, *Weapons of Math Destruction*, Penguin Books, New York, trad. it. *Armi di distruzione matematica*, Bompiani, Milano, Bompiani, 2017; D.K. CITRON, F. PASQUALE, *The scored society: Due process for automated predictions*, in *Wash. L. Rev.*, 2014, 1–34.

⁹³ A. FINE, e al., *Public Perceptions of Judges’ Use of AI Tools in Courtroom Decision-Making: An Examination of Legitimacy, Fairness, Trust, and Procedural Justice*, in *Behavioral Sciences*, 4, 2025, 1–21

⁹⁴ Qui adattiamo la triade già individuate da J. BURRELL, *How the machine ‘thinks’: Understanding opacity in machine learning algorithms*, cit.

⁹⁵ L’opacità istituzionale si verifica quando l’accesso alle informazioni rilevanti (ad esempio, dettagli del modello, dati di addestramento, valutazione) è limitato da considerazioni di segreto commerciale, sicurezza, privacy o anti-gaming (ad esempio, strumenti proprietari di valutazione del rischio utilizzati nella giustizia penale). B. VAASSEN, *AI, Opacity, and Personal Autonomy*, cit.; D. PURVES, J. DAVIS, *Public Trust, Institutional Legitimacy, and the Use of Algorithms in Criminal Justice*, in *Public Affairs Quarterly*, 2, 2022, 136–162; R. WEXLER, *Life, liberty, and trade secrets: Intellectual property in the criminal justice system*, cit.

⁹⁶ L’opacità legata alle competenze o alla comunicazione si verifica quando le informazioni sono tecnicamente disponibili ma non utilizzabili dai principali soggetti interessati a causa di barriere legate alle competenze, al tempo o alla traduzione; ciò motiva il lavoro sull’alfabetizzazione algoritmica e sulle spiegazioni adeguate per i soggetti interessati. V. CHIAO, *Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice*, in *International Journal of Law in Context*, 2, 2019, 126–139.

⁹⁷ L’opacità tecnico-epistemica deriva dalla complessità del modello e dalla portata delle pipeline socio-tecniche, che possono limitare un esame significativo anche in caso di accesso; ciò motiva l’interpretabilità fin dalla progettazione, la documentazione strutturata e i quadri di valutazione per la qualità delle spiegazioni. S. CHESTERMAN, *Through a glass, darkly: artificial intelligence and the problem of opacity*, in *The American Journal of Comparative Law*, 2, 2021, 271–294

⁹⁸ R. WEXLER, *Life, liberty, and trade secrets: Intellectual property in the criminal justice system*, cit., 1343.

⁹⁹ Insieme alla letteratura precedente, si v. anche K. YEUNG, A. HARKENS, *How do ‘technical’ design-choices made when building algorithmic decision-making tools for criminal justice authorities create constitutional dangers? (Part I)*, in *Public law*, Apr, 2023, 265–286

¹⁰⁰ Revisioni da parte di esperti, test contraddittori, esperti nominati dal tribunale e decisioni attente su quali soggetti (giudici, avvocati, parti, giurati) devono comprendere cosa e a quale livello di dettaglio. Nei processi con giuria, ad

Infine, l'opacità tecnica o epistemica (intrinseca) solleva una questione diversa: che cosa può richiedere la legge quando un sistema non può essere reso completamente trasparente in termini comprensibili all'uomo? In questo caso, la dottrina giuridica già contiene strategie per affrontare l'opacità del processo decisionale umano. I tribunali spesso richiedono una motivazione, ovvero una spiegazione a posteriori dei motivi delle decisioni (ad esempio, nelle sentenze giudiziarie o nelle decisioni amministrative). Ciò suggerisce un approccio analogo per i sistemi di IA: piuttosto che insistere sulla piena interpretabilità del modello stesso, i legislatori e i giudici potrebbero richiedere spiegazioni e registrazioni contestabili e specifiche per ogni decisione (input considerati, fattori che hanno determinato il risultato in senso probatorio, limiti di incertezza e risultati di valutazione documentati), insieme a diritti di controllo e rimedi quando le spiegazioni risultano inadeguate¹⁰¹.

Tornando alla questione affrontata all'inizio di quest'analisi sull'opacità, queste considerazioni influiscono direttamente sul modo in cui si dovrebbe concepire il diritto alla spiegazione delle decisioni giudiziarie assistite dall'intelligenza artificiale, ovvero il diritto delle persone interessate di ottenere informazioni significative su come è stata presa la decisione, le quali siano sufficienti a consentirne la comprensione, la contestazione e la revisione. La definizione di un diritto alla spiegazione socialmente vantaggioso richiede attenzione a (1) il tipo di decisione giudiziaria in questione (ad esempio: condanna, libertà vigilata, responsabilità civile), (2) il tipo di spiegazione fornita (ad esempio, motivazioni, prove controfattuali, documentazione, verifiche), (3) i probabili effetti sulla capacità delle persone di contestare i risultati e (4) i più ampi benefici e rischi sistemici della divulgazione (tra cui accuratezza, legittimità, uguaglianza e fiducia istituzionale)¹⁰².

Le obiezioni più comuni al riconoscimento del diritto alla spiegazione – segreto commerciale e preoccupazioni che i cittadini possano “aggirare il sistema” – invocano principalmente l'opacità istituzionale (intenzionale). La protezione del segreto commerciale viene talvolta invocata per giustificare l'occultamento di alcuni o tutti gli elementi di un sistema decisionale algoritmico, compresi le specifiche del modello, i dati di addestramento, i risultati di convalida e persino le informazioni di base necessarie per contestare un risultato¹⁰³.

Soprattutto nei procedimenti penali, tale omissione è stata criticata in quanto difficile da conciliare con le tradizioni di equità procedurale e con la capacità dell'imputato di verificare le prove contrarie. Nel giudicare questi conflitti, i giudici dovrebbero trattare il segreto commerciale non come un ostacolo automatico, ma come un interesse da bilanciare con l'interesse pubblico alla trasparenza, alla responsabilità e all'efficace esercizio dei diritti¹⁰⁴. I tribunali possono anche mitigare le legittime preoccupazioni in materia di riservatezza attraverso strumenti procedurali già usati altrove. Ad esempio, possono consentire la revisione a porte chiuse, nominare esperti neutrali o emettere

esempio, i giudici possono mitigare l'opacità dovuta alla competenza assicurandosi che le parti abbiano opportunità significative di contestare le prove algoritmiche e che i giurati ricevano strumenti - attraverso istruzioni, testimonianze di esperti verificati e limitazioni alle affermazioni esagerate - per interpretare correttamente i risultati. A. ROTH, *Trial by machine*, in *Geo. LJ*, 5, 2015, 1245–1306.

¹⁰¹ Questo modello di “motivazione più contestabilità” è compatibile con alcuni sistemi *black box*, anche se andrebbe abbinato a misure di salvaguardia contro l'eccessiva dipendenza da motivazioni post hoc potenzialmente fuorvianti. Z.C. LIPTON, *The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery*, in *Queue*, 3, 2018, 31–57.

¹⁰² In questa prospettiva, il diritto alla spiegazione dovrebbe essere calibrato in base al contesto, anziché concepito come una regola di divulgazione unica e uniforme. Si v. in questo senso il caso C-634/21, SCHUFA Holding (Scoring), 7 dicembre 2023.

¹⁰³ Tema già ampiamente noto fin dalla sentenza approvata dalla Corte distrettuale dell'Aia (Paesi bassi) del 5 febbraio 2020 (ECLI:NL:RBDHA:2020:1878) nel caso relativo al sistema SyRI.

¹⁰⁴ Come afferma il Consiglio di Stato, Sez. VI, 8 aprile 2019, n. 2270.

ordinanze cautelari che limitano l'uso e la diffusione di informazioni sensibili al di là delle esigenze del procedimento¹⁰⁵.

La “divulgazione” non è una questione di tutto o niente¹⁰⁶. Essa dipende dal contesto decisionale e dall'importanza delle preoccupazioni relative ai segreti commerciali o alle pratiche anti-*gaming*¹⁰⁷.

Inoltre, l'utilità della divulgazione dipende dal pubblico e dallo scopo: per la maggior parte delle persone interessate, le informazioni più significative dal punto di vista pratico saranno le variabili prese in considerazione da un sistema (e le loro definizioni operative), mentre gli esperti possono utilizzare le specifiche del modello, il codice e la documentazione di formazione per verificare la validità e la completezza¹⁰⁸.

Per questi motivi, i dibattiti sulla trasparenza dell'IA nel processo decisionale giudiziario dovrebbero andare oltre la questione binaria se divulgare o meno e specificare quali informazioni devono essere divulgate, a chi e con quali garanzie procedurali, per rendere significativo il diritto alla spiegazione.

2.3. Bias, opacità e diritto antidiscriminatorio

L'esame dei casi giudiziari e dell'abbondante letteratura giuridica – qui solo in parte riportata – relativa ai pregiudizi e all'opacità intrinseca all'uso degli algoritmi nei processi ci ha portati a validare l'ipotesi iniziale, emersa alcuni anni fa nell'esame del caso *Loomis* (COMPAS), secondo cui i processi decisionali algoritmici possono produrre discriminazioni anche notevoli¹⁰⁹.

Come i decisori umani, i sistemi algoritmici possono generare effetti discriminatori anche in assenza di intenzione di discriminare. Uno dei risultati più interessanti dell'esame

¹⁰⁵ Un regime di diritto alla spiegazione può essere calibrato attraverso una serie di divulgazioni, che vanno dalla provenienza dei dati e dalle pratiche di conservazione, alle prove di convalida, e alle definizioni operative delle caratteristiche, alle specifiche del modello, all'importanza e al peso delle caratteristiche, ove significativo, alle variabili di *output* e alle soglie, nonché ai registri specifici del caso sufficienti per la contestazione, implementati attraverso strumenti familiari di tutela della riservatezza, quali il deposito sotto sigillo, la revisione a porte chiuse e le ordinanze cautelari che limitano la diffusione oltre le esigenze del procedimento. Se strutturati in modo adeguato, questi meccanismi possono conciliare gli interessi privati con i vantaggi della trasparenza, garantendo che le informazioni necessarie per la responsabilità e una contestazione significativa siano disponibili alle parti e al tribunale, senza trasformare i procedimenti giudiziari in un veicolo per la divulgazione pubblica senza restrizioni. Si v. in questo senso il caso C-203/22, CK c. Dun & Bradstreet Austria GmbH, 12 settembre 2024.

¹⁰⁶ K. FOSS-SOLBREKK, *Searchlights across the black box: Trade secrecy versus access to information*, in *Computer Law & Security Review*, 2023, 1–17.

¹⁰⁷ Anche le obiezioni relative al “gaming” dovrebbero essere valutate con specificità. La manipolazione dannosa richiede in genere una conoscenza dettagliata di quali input siano rilevanti e di come vengano combinati, quindi è spesso possibile fornire informazioni rilevanti per la contestazione attraverso una divulgazione selettiva, trattenendo i dettagli che potrebbero consentire l'evasione o l'inganno; un approccio che si riflette nelle pratiche di valutazione del merito creditizio, che divulgano i fattori chiave senza rivelare completamente i modelli di valutazione. I.N. COFONE, K.J. STRANDBURG, *Strategic Games and Algorithmic Secrecy*, in *McGill Law Journal*, 4, 2020, 623–663.

¹⁰⁸ Poiché l'accesso a livello di codice tende a incontrare la maggiore resistenza ed è raramente comprensibile ai non addetti ai lavori, il suo valore principale risiede nel consentire una verifica indipendente piuttosto che facilitare la comprensione individuale; di conseguenza, i regimi di divulgazione dovrebbero essere calibrati in base al contesto e non devono necessariamente essere tutto o niente. In contesti come COMPAS, gli imputati possono trarre il massimo vantaggio dalla divulgazione delle caratteristiche (ad esempio, accuse, precedenti penali, abuso di sostanze, istruzione, occupazione), mentre i tribunali possono spesso richiedere la divulgazione delle caratteristiche - o delle fonti dei dati e delle pratiche di costruzione da cui derivano - senza imporre la divulgazione di ponderazioni, soglie o codici proprietari; al contrario, giornalisti, ricercatori e organizzazioni di interesse pubblico potrebbero richiedere un accesso più approfondito per individuare gli effetti sistemici, e anche la sola divulgazione delle fonti dei dati di addestramento può migliorare concretamente la responsabilità, consentendo il controllo dei pregiudizi di gruppo incorporati. Si v. sul punto J. LARSON, S. MATTU, L. KIRCHNER, J. ANGWIN, *How we analyzed the COMPAS recidivism algorithm*, cit.

¹⁰⁹ Si v. a questo proposito l'interessante esame che compie K. CRAWFORD, *Né intelligente né artificiale. Il lato oscuro dell'IA*, il Mulino, Bologna, 2021, 146 ss.

di tali casi è che la legislazione antidiscriminatoria risulterebbe inadeguata ad affrontare questi eventi,¹¹⁰ almeno per due ragioni.

In *primo* luogo, l'attenzione dottrinale alle leggi antidiscriminazione, insieme al carattere prevalentemente *ex post* dei rimedi giuridici, spesso fatica a contrastare la discriminazione sistemica, proprio il tipo di disuguaglianza strutturale che i sistemi algoritmici possono catturare, riprodurre e talvolta intensificare¹¹¹.

In *secondo* luogo, anche quando un modello non utilizza esplicitamente dati di soggetti vulnerabili, può assorbire pregiudizi dai dati di addestramento e da modelli sociali e storici radicati, e può basarsi su variabili apparentemente neutre che fungono da *proxy* per tratti protetti, producendo forme di discriminazione indiretta difficili da identificare e porre rimedio¹¹².

Laddove il trattamento iniquo può essere ricondotto a un processo decisionale che tratta i gruppi in modo diverso, l'adeguatezza alle categorie antidiscriminatorie familiari (in particolare la discriminazione diretta) è relativamente semplice. Al contrario, il pregiudizio del campione e il rafforzamento delle disuguaglianze sociali di fondo spesso non si adattano perfettamente alle norme esistenti che impediscono le discriminazioni indirette¹¹³, rendendole più difficili da diagnosticare sia per la dottrina sia in pratica.

Quest'esame ha evidenziato che le discriminazioni algoritmiche non possono essere affrontate semplicemente trasferendo le dottrine sviluppate per il mondo analogico all'interno dei contesti in cui sono utilizzati sistemi computazionali.

Dato che gli effetti discriminatori algoritmici derivano da meccanismi basati sui dati¹¹⁴, è spesso necessario un approccio preventivo¹¹⁵. Le norme sulla protezione dei dati e sulla privacy possono contribuire a tale approccio regolando la raccolta, la qualità, l'uso e il riutilizzo dei dati personali su cui vengono addestrati e implementati i modelli: principi quali la minimizzazione e l'accuratezza dei dati, la "protezione dei dati fin dalla progettazione e per impostazione predefinita" e strumenti procedurali come le valutazioni d'impatto sulla protezione dei dati possono limitare la produzione di bias e richiedere alle organizzazioni di identificare e mitigare i rischi per i diritti e le libertà, comprese le discriminazioni ingiuste. Allo stesso tempo, ciò non rende superflua la legislazione antidiscriminatoria, ma evidenzia piuttosto una divisione dei compiti: i rimedi antidiscriminatori tendono ad agire *ex post* e a lottare contro modelli diffusi e sistemici, mentre la governance dei dati può talvolta intervenire a monte, modellando gli input informativi e le pratiche che generano output discriminatori¹¹⁶.

Queste lacune giuridiche rendono particolarmente rilevante il ruolo del potere giudiziario. I giudici non sono solo arbitri di ammissibilità e decisori del merito, ma anche guardiani istituzionali che devono decidere quando e come affidarsi ai risultati algoritmici

¹¹⁰ J. ADAMS-PRASSL, e al., *Directly Discriminatory Algorithms*, in *The Modern Law Review*, 1, 2023, 144–175; C. NARDOCCI, *Algoritmi, eguaglianza, discriminazione: le sfide dell'intelligenza artificiale*, Giappichelli, Torino, 2025.

¹¹¹ In letteratura questo tema è conosciuto da molto: S. BAROCAS, A.D. SELBST, *Big data's disparate impact*, in *Calif. L. Rev.*, 3, 2016, 671–732.

¹¹² A.E. PRINCE, D. SCHWARCZ, *Proxy discrimination in the age of artificial intelligence and big data*, in *Iowa L. Rev.*, 3, 2019, 1257–1318.

¹¹³ S. WACHTER, e al., *Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI*, in *Computer Law & Security Review*, 2021, 1–72; V. EUBANKS, *Automating inequality: How high-tech tools profile, police, and punish the poor*, St. Martin's Press, New York, 2018.

¹¹⁴ Per esempio campioni di addestramento distorti, errori di misurazione e l'uso di proxy apparentemente neutri che riproducono disuguaglianze strutturali.

¹¹⁵ Come evidenziato da S. BAROCAS, A.D. SELBST, *Big data's disparate impact*, cit.

¹¹⁶ P.T. KIM, *Data-driven discrimination at work*, in *Wm. & Mary L. Rev.*, 3, 2016, 857–936; S. WACHTER, B. MITTELSTADT, L. FLORIDI, *Why a right to explanation of automated decision-making does not exist in the general data protection regulation*, cit.

nelle sentenze¹¹⁷. Tale responsabilità richiede ai tribunali di comprendere le principali fonti di pregiudizio algoritmico e le loro implicazioni sia per l'accuratezza sia per la discriminazione, di esaminare attentamente le condizioni in cui uno strumento di IA è stato sviluppato e convalidato e, laddove gli strumenti automatizzati abbiano un impatto significativo sulle parti, di garantire garanzie procedurali quali la possibilità di contestare la base del risultato e di ottenere una revisione significativa¹¹⁸.

Come vedremo tra poco, l'IA generativa ha in parte risolto alcuni di questi problemi, ma ne ha creati altri prima non conosciuti. Oggi, infatti, si parla sempre più di "opacità sistemica", la quale emergerebbe dall'interazione tra vari modelli di IA generativa, le API e i database. Ciò crea una sorta di scatola nera "emergente" in cui, anche se i singoli componenti sono comprensibili, la logica collettiva dell'intero *stack* tecnologico diventa imprevedibile e impossibile da mappare. Nel loro insieme, queste dimensioni dimostrano che l'opacità non è soltanto un problema tecnico, ma una sfida complessa che coinvolge dati, legislazione, istruzione e, non secondariamente, l'architettura dei sistemi.

3. Gli anni venti: l'arrivo dell'IA generativa (GenAI)

Come già anticipato, passiamo ora all'esame delle nuove tecnologie di rilievo nel diritto e mettiamo come punto di avvio del nostro discorso il mese di novembre del 2022, quando è apparso l'LLM conversazionale "ChatGPT"¹¹⁹.

Gli LLM (Large Language Models) sono una sottocategoria della GenAI¹²⁰ che sfrutta le capacità dei modelli di deep learning avanzati per generare testo. Gli LLM sono modelli di

¹¹⁷ E. LONGO, *Giustizia digitale e Costituzione: riflessioni sulla trasformazione tecnica della funzione giurisdizionale*, cit.

¹¹⁸ In fondo queste sono le conclusioni sia del caso *Loomis* sia del caso *SyRI* nei quali i giudici hanno dovuto decidere quale peso attribuire ai risultati di una *black box* e se questi violassero il giusto processo. Il tema del giusto processo è un aspetto essenziale in questo ambito come evidenziato da D.K. CITRON, *Technological due process*, in *Wash. UL Rev.*, 6, 2008, 1249–1314 e poi da F. PASQUALE, *Inalienable Due Process in an Age of AI: Limiting the Contractual Creep toward Automated Adjudication*, cit., 42 ss.

¹¹⁹ ChatGPT rappresenta il primo LLM conversazionale a raggiungere una diffusione di massa, grazie al suo rilascio pubblico da parte di OpenAI nel novembre 2022. Il suo fattore di successo decisivo risiede nell'essere addestrato specificamente per interagire con gli utenti in linguaggio naturale, producendo risposte fluide e contestualmente appropriate, attraverso una tecnica nota come *Reinforcement Learning from Human Feedback* (RLHF). OpenAI è un'azienda fondata nel 2015 come no-profit finanziata da molte delle principali aziende "big tech" americane per studiare le applicazioni positive dell'IA. È diventata un'azienda commerciale nel 2019. Nel 2023 è stata acquisita da Microsoft per circa 10 miliardi di dollari. OpenAI, *ChatGPT: Optimizing Language Models for Dialogue*, 30 novembre 2022 (reperibile in: <https://openai.com/blog/chatgpt>); P.P. RAY, *ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope*, in *Internet of Things and Cyber-Physical Systems*, 2023, 121–154; OpenAI, *Presentazione di ChatGPT: Optimizing Language Models for Dialogue*, OPEN AI: BLOG (30 novembre 2022), <https://openai.com/blog/chatgpt> [<https://perma.cc/8QWZ-7NKY>].

¹²⁰ L'IA generativa, che è anche fondazionale, è un tipo di IA che attinge alla progettazione di reti neurali profonde e agli sviluppi dell'apprendimento automatico, individuando modelli nei dati esistenti per generare output nuovi e unici, quali testi, video, immagini, audio o modelli 3D. Secondo la formulazione della Sezione 3 dell'Ordine Esecutivo degli Stati Uniti dell'ottobre 2023, l'IA generativa "indica la classe di modelli di IA che emulano la struttura e le caratteristiche dei dati di input al fine di generare contenuti sintetici derivati". US President Joe Biden, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, Executive Order 14110, 2023 (reperibile in: <https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence>). Si veda anche T. MELHAM, *Generative AI*, in *The Cambridge Handbook of Generative AI and the Law*, a cura di M. ZOU, et al., Cambridge University Press, Cambridge, 2025, 3–10; Congressional Research Service, *Generative artificial intelligence and data privacy: A primer*, No. R47569, 23/5/2023 (reperibile in: <https://crsreports.congress.gov/product/pdf/R/R47569>); R. BOMMASANI, e al., *On the opportunities and risks of foundation models*, in *arXiv preprint arXiv:2108.07258*, 2021, (reperibile in <https://arxiv.org/pdf/2108.07258/pdf>); I. TRIGUERO, e al., *General Purpose Artificial Intelligence Systems (GPAIS): Properties, definition, taxonomy, societal implications and responsible governance*, online first, 2024, 1–16.

apprendimento automatico addestrati su enormi quantità di dati testuali, con l'obiettivo di generare testi statisticamente coerenti con il linguaggio umano, producendo risposte in linguaggio naturale¹²¹: testi scritti, espressioni significative, contenuti articolati e argomentazioni strutturate. Nel panorama dell'IA, gli studiosi associano l'LLM anche al sottocampo noto come "Natural Language Processing" (NLP)¹²².

La storia degli LLM risale almeno al 2018, con l'introduzione di BERT da parte di Google e di GPT da parte di OpenAI, entrambi basati sull'architettura Transformer introdotta da un gruppo di ricercatori di Google nel 2017¹²³.

Dal novembre 2022, modelli linguistici di grandi dimensioni e strumenti analoghi hanno conosciuto una diffusione straordinaria, diventando strumenti sempre più presenti nella vita quotidiana e professionale. Le loro capacità si sono ampliate e affinate in modo significativo nel breve volgere di pochi anni, con il rilascio di modelli sempre più potenti da parte di diversi operatori — tra cui Google (Gemini), Meta (LLaMA), Anthropic (Claude) e altri — che si affiancano ai modelli di OpenAI (GPT-5.2 e versioni successive).

Pertanto, concetti come "comprensione" e "intelligenza" dovrebbero essere considerati come "metafore funzionali"¹²⁴; sarebbe fuorviante caratterizzare questi sistemi come "intelligenti" nello stesso senso in cui lo sono gli esseri umani cognitivamente attivi. Fondamentalmente, questi grandi modelli linguistici (LLM) rimangono macchine di pattern-matching statistico molto avanzate, nonostante i loro risultati impressionanti.¹²⁵

Il settore giuridico-legale è stato tra i primi in cui è apparso evidente il potenziale impatto di queste tecnologie¹²⁶. La ragione è strutturale: la stragrande maggioranza dei documenti giuridici — sentenze, contratti, atti processuali, pareri legali, normative — assume forma scritta. Anche nei casi in cui gli atti processuali vengono compiuti oralmente, è comunque necessario conservarne traccia scritta nei verbali d'udienza. Questa natura intrinsecamente testuale del diritto rende il settore legale particolarmente

¹²¹ Sono chiamati "grandi" perché sono addestrati su un ampio set di dati e sono in grado di generare testi altamente realistici e coerenti, diversi a ogni richiesta.

¹²² Il termine "linguaggio naturale" ha un significato specifico nell'informatica, in quanto si riferisce alle lingue quotidiane che le persone usano per comunicare, tra cui l'inglese, il francese e lo spagnolo. Questo termine si contrappone all'idea di "linguaggi formali" dell'informatica, che sono "linguaggi tecnici" precisi e matematicamente strutturati, come Python e JavaScript, usati per programmare i computer e per altre attività. I linguaggi formali sono progettati per un'interpretazione meccanica non ambigua, con conseguente espressività limitata, ma elaborazione affidabile da parte dei computer.

L'NLP è l'area di ricerca che si concentra sulla creazione di sistemi di intelligenza artificiale in grado di comprendere e generare testi e linguaggio umano. Fino al 2020, i sistemi NLP non erano in grado di interpretare un testo scritto a un livello nemmeno lontanamente paragonabile a quello di un essere umano. I sistemi NLP di quel periodo analizzavano principalmente la comunicazione umana utilizzando semplici schemi di parole e dati. Pertanto, sebbene questi (vecchi) sistemi di NLP fossero in grado di identificare in modo affidabile parole chiave e modelli di testo all'interno di documenti come le e-mail, faticavano a comprendere realmente il contenuto di qualsiasi documento in linguaggio naturale. (N. ALETRAS, e al., *Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective*, in *PeerJ Computer Science*, 2, 2016, 2–19). Queste limitazioni si sono protratte per decenni, portando molti ricercatori a considerare il potenziale dei computer di comprendere efficacemente il linguaggio umano come un obiettivo di ricerca relativamente lontano. I modelli NLP più avanzati venivano definiti nient'altro che "pappagalli stocastici" perché davano l'illusione di comprendere il linguaggio naturale, quando in realtà la maggior parte del testo apparentemente coerente generato dai sistemi dell'epoca era ottenuta rimescolando versioni di frasi esistenti scritte in precedenza su Internet. Si v. E.M. BENDER, e al., *On the Dangers of Stochastic Parrots*, in Canada, 2021 (reperibile in: <https://dx.doi.org/10.1145/3442188.3445922>)

¹²³ A. VASWANI, e al., *Attention Is All You Need*, Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS), 2017 (reperibile in: <https://arxiv.org/abs/1706.03762>).

¹²⁴ H. SURDEN, *ChatGPT, Artificial Intelligence (AI) Large Language Models, and Law*, in *Fordham Law Review*, 5, 2024, 1942–1972.

¹²⁵ R.T. MCCOY, e al., *Embers of Autoregression: Understanding Large Language Models Through the Problem They are Trained to Solve*, 2023 (reperibile in: <https://arxiv.org/abs/2309.13638>).

¹²⁶ J.H. CHOI, e al., *Lawyering in the age of artificial intelligence*, 2023, Minnesota Legal Studies Research Paper No. 23-31 (reperibile in: <https://ssrn.com/abstract=4626276>).

permeabile alle applicazioni della GenAI basate sull'elaborazione del linguaggio naturale e sulle correlazioni statistiche tra testi¹²⁷.

Ciò avviene grazie a un algoritmo GPT (Generative Pretrained Transformer)¹²⁸ che è in grado di leggere le parole incorporate (Word Vector) anche grazie a “Instruction Finetuning” e “Reinforcement Learning”¹²⁹.

Tutti ricorderanno che già nel marzo 2023 un giudice colombiano ha emesso una sentenza raccontando del proprio “dialogo” con ChatGPT¹³⁰.

Quell'episodio ha reso concreta — e forse inevitabile — la prospettiva di delegare porzioni del ragionamento giuridico a un sistema di intelligenza artificiale generativa. Non si è trattato di un caso isolato: negli anni successivi si sono moltiplicati episodi analoghi, alcuni ancora più problematici del primo¹³¹. A confermare questa tendenza, una recente indagine dell'UNESCO ha rilevato che il 44% degli operatori giudiziari — giudici, pubblici ministeri e avvocati — sta già integrando strumenti come ChatGPT nella propria attività professionale. Dato ancor più preoccupante: solo il 9% di questi professionisti ha ricevuto una formazione istituzionale adeguata o dispone di linee guida formali per l'utilizzo di tali tecnologie¹³².

Alla luce della letteratura esaminata nella sezione precedente, questi casi sembrano collocarsi lungo un *continuum* tra due approcci distinti. Il primo è un uso ausiliario: il professionista si avvale dell'algoritmo per identificare rilevanti precedenti, selezionare le norme applicabili o orientarsi nella giurisprudenza. Il secondo è un uso generativo: i fatti di causa vengono sottoposti al sistema, che — addestrato su decisioni pregresse e normativa vigente — produce direttamente il testo della sentenza o della motivazione.

È ovvio che in nessuno di questi scenari si può sostenere che ChatGPT “decida” il caso. Gli LLM non possiedono né la comprensione del testo che generano né la volontà cognitiva necessaria a formulare un giudizio. Eppure, proprio questo solleva una domanda che non può essere elusa: qual è il ruolo di un sistema che genera un testo poi utilizzato — in modi che non sempre è possibile ricostruire — da un giudice nella fase decisoria?

¹²⁷ Come osservato in letteratura, gli LLM sono in grado di analizzare e modificare documenti legali, estrarne elementi chiave, generare bozze di atti, individuare casi simili e fornire supporto alla ricerca giuridica — funzioni che ampliano notevolmente le possibilità operative per giuristi, avvocati e magistrati, pur nella consapevolezza dei limiti e dei rischi che tali tecnologie comportano. S.S. TU, e al., *Artificial intelligence: Legal reasoning, legal research and legal writing*, in *Minn. JL Sci. & Tech.*, Symposium Issue, 2023, 105–126; Z. HOU, e al., *Large language models meet legal artificial intelligence: A survey*, in *arXiv preprint arXiv:2509.09969*, 2025, (reperibile in: <https://arxiv.org/abs/2509.09969>); H. SURDEN, *ChatGPT, Artificial Intelligence (AI) Large Language Models, and Law*, cit..

¹²⁸ I GPT sono pre-addestrati su una grande quantità di dati per cogliere le sfumature del linguaggio, delle immagini e dei suoni e generare testi, immagini o suoni coerenti e consapevoli del contesto.

¹²⁹ S. FEUERRIEGEL, e al., *Generative AI*, in *Business & Information Systems Engineering*, 1, 2024, 111–126.

¹³⁰ Nel gennaio 2023, in Colombia si è verificata una controversia quando un giudice ha rivelato di aver utilizzato ChatGPT per determinare se l'assicurazione di un bambino autistico dovesse coprire tutte le spese relative alle cure mediche. Dieci giorni dopo, sempre in Colombia, un magistrato ha emesso un'ordinanza con l'assistenza di ChatGPT per decidere di condurre un processo nel metaverso. Inoltre, a fine marzo 2023, un giudice in Perù e un magistrato in Messico hanno ammesso di aver utilizzato ChatGPT di OpenAI per supportare una decisione di secondo grado e per illustrare argomenti in un'udienza. In seguito al caso *Mata vs. Avianca Airlines, Inc.* in cui un avvocato ha presentato citazioni e casi falsificati generati da ChatGPT a un tribunale statunitense, l'importanza di linee guida per un uso responsabile è diventata ancora più evidente (*Mata v. Avianca, Inc.*, 1:22-cv-01461, disponibile su: <https://www.courtlistener.com/docket/63107798/mata-v-avianca-inc/>). In generale, su questi casi, si veda R.

SCHWORER, *Ethical Implications of Utilizing Generative Artificial Intelligence in the Legal Profession: A Cautionary Note on Potential Model Rules of Professional Conduct Violations*, in *Lincoln Memorial University Law Review Archive*, 2, 2024, 51–75; R. PERONA, Y.C. DE LA ROSA, *Unveiling AI in the courtroom: exploring ChatGPT's impact on judicial decision-making through a pilot Colombian case study*, in *AI & SOCIETY*, online first, 2024, 1–8. La storia e il problema del primo giudice noto che utilizza ChatGPT sono illustrati da J.D. GUTIÉRREZ, *ChatGPT in Colombian Courts: Why we need to have a conversation about the digital literacy of the judiciary*, cit..

¹³¹ R. SCHWORER, op. ult. cit.; H. SURDEN, *ChatGPT, Artificial Intelligence (AI) Large Language Models, and Law*, cit..

¹³² UNESCO, *Global Judges' Initiative: Survey on the use of AI systems by judicial operators*, CI/DIT/2024/JI/01 Rev., 2024 (reperibile in: <https://unesdoc.unesco.org/ark:/48223/pf0000389786>)

3.1. Gli usi dei LLM nel campo del diritto, con una prima impressione su di essi

La GenAI rappresenta oggi un'industria colossale¹³³. Questa innovativa tecnologia digitale ha un immenso potenziale per aumentare in modo significativo l'efficienza e la produttività in molti settori, snellendo processi e decisioni specifiche¹³⁴.

Oggi, come già detto, i sistemi LLM, come ChatGPT, sono in grado di svolgere nuovi compiti in ambito giuridico-legale: (1) produrre documenti legali ragionevoli e di qualità, come contratti, ricorsi e sentenze; (2) eseguire analisi legali di base; (3) rispondere a domande su documenti legali, come leggi, pareri legali o contratti¹³⁵.

Gli LLM sono in grado di identificare argomentazioni o riassumere documenti molto lunghi e di generare testi, tra cui argomentazioni e ricerche giuridiche¹³⁶, prevedendo il testo appropriato da far seguire a un dato input, utilizzando modelli appresi da grandi insiemi di dati¹³⁷.

Questi strumenti stanno cambiando le carte in tavola in molti settori. Sono molto più rivoluzionari dei motori di ricerca web apparsi trent'anni fa, che pure hanno sconvolto la diffusione e la fruibilità di leggi e sentenze come mai prima d'ora. I nuovi chatbot offrono molto di più, poiché possono assumere il ruolo delle persone che un tempo redigevano documenti, non solo quello di assistenti personali o di professionisti legali.

A metà del 2024, Adam Unikowsky ha condotto un esperimento sistematico per misurare le capacità dei modelli generativi in ambito giudiziario (“Anthropic Claude Opus 3”), dimostrando che l'IA non si limita ad assistere il giudice nelle fasi ancillari del lavoro — come la ricerca della giurisprudenza pertinente — ma è in grado di redigere autonomamente bozze di pareri motivati su casi di appello, avvicinandosi per qualità e sofisticatezza alle decisioni delle corti di legittimità¹³⁸. L'esperimento ha mostrato che i modelli sono capaci di formulare nuovi test legali, applicarli ai fatti del caso e persino ai confini ipotetici della norma, e di proporre criteri e motivazioni originali che vanno oltre quanto suggerito dalle parti o consolidato nella giurisprudenza precedente.

¹³³ EDITORIAL, *Large, creative AI models will transform lives and labour markets*, *The Economist*, 22/4/2023 (reperibile in: <https://www.economist.com/interactive/science-and-technology/2023/04/22/large-creative-ai-models-will-transform-how-we-live-and-work>).

¹³⁴ J. NELSON, *The Other 'LLM': Large Language Models and the Future of Legal Education*, in *European Journal of Legal Education*, 1, 2024, 127–155.

¹³⁵ *Infra multis v. D. SCHWARCZ, J.H. CHOI, AI Tools for Lawyers: A Practical Guide*, in *Minnesota Law Review Headnotes*, 1, 2023, 1–39; Y. RAZMETAeva, *Artificial intelligence and the end of justice*, in *Biolaw Journal*, 1, 2024, 345–365; J.D. GUTIÉRREZ, *Critical appraisal of large language models in judicial decision-making*, in *Handbook on Public Policy and Artificial Intelligence*, a cura di R. PAUL, et al., Edward Elgar Publishing, London, <https://dx.doi.org/10.4337/9781803922171.00033>, 2024, 323–338; S. FRALEY, *Digital Justice: Shining Future or Illusory Chimera?*, in *Pólemos*, 1, 2020, 47–55, 47; P. RUBIM BORGES FORTES, *Paths to Digital Justice: Judicial Robots, Algorithmic Decision-Making, and Due Process*, in *Asian Journal of Law and Society*, 2020, 1–17, 1.

¹³⁶ Mentre alcuni chatbot LLM operano all'interno di un universo chiuso di informazioni, altri sono aperti e hanno un accesso più ampio ai dati, ad esempio tramite plug-in web o connessioni Internet. È già possibile sfruttare gli LLM su specifici insiemi di dati o corpi di testo e focalizzarli su compiti ristretti. Attraverso queste applicazioni, si possono creare enciclopedie personali e persino veri e propri database per addestrare il proprio chatbot.

¹³⁷ Ad esempio, il GPT-4 è in grado (anche se a volte in modo imperfetto) di fare ragionamenti giuridici su leggi e fatti, di analizzare o generare contratti, di riassumere casi giuridici, di redigere brevetti, di scrivere mozioni e di rispondere a domande su pareri o documenti legali. Cfr. H. SURDEN, *ChatGPT, Artificial Intelligence (AI) Large Language Models, and Law*, cit..

¹³⁸ Gli esperimenti sono descritti in dettaglio in Adam Unikowsky, “In AI we trust”, 8 giugno 2024, <https://adamunikowsky.substack.com/p/in-ai-we-trust> consultato il 4 marzo 2026 e “In AI we trust, part II”, 16 giugno 2024, <https://adamunikowsky.substack.com/p/in-ai-we-trust>, consultato il 4 marzo 2026. Siamo venuti a conoscenza dell'esperimento consultando il saggio di R. DIAB, *The evolving role of AI in legal judgment*, in *Law, Innovation and Technology*, 2026, 1–32.

Dal punto di vista scientifico, il salto concettuale è notevole. La traiettoria che collega il primo uso di ChatGPT documentato in Colombia all'esperimento di Unikowsky descrive una transizione da un uso accidentale e passivo dell'IA come strumento di supporto argomentativo a un utilizzo strutturato e attivo come agente del ragionamento giuridico. In termini cognitivi, si tratta del passaggio da un sistema assistivo, di *retrieval* (recupero di informazioni preesistenti), a un sistema generativo, di *reasoning* (generazione di nuove inferenze normative). Questa distinzione è cruciale: se nel primo caso l'IA amplifica la memoria del giudice, nel secondo ne estende – o ne simula – la capacità deliberativa¹³⁹.

Allora, la domanda scientifica che ne emerge ci riporta al punto di partenza di questo saggio: non ci si chiede più se l'IA possa supportare la giustizia, ma in quale misura la produzione di motivazioni giuridiche coerenti e originali da parte di un sistema computazionale possa essere considerata analoga – o addirittura equivalente, sotto certi profili funzionali – al ragionamento giuridico umano. La risposta a questa domanda non appartiene solo agli studiosi di diritto, ma richiama anche la filosofia della mente, la teoria dell'argomentazione e la scienza cognitiva, aprendo un terreno di ricerca interdisciplinare ancora in gran parte inesplorato¹⁴⁰.

3.2. Dai problemi tecnici alle ricadute costituzionali

Per comprendere le implicazioni costituzionali della delega del processo decisionale a un LLM, è innanzitutto importante approfondire le preoccupazioni tecniche e i rischi potenziali associati all'uso di questi strumenti. Questi rischi comportano la possibilità che questi sistemi generino risultati errati, noti come “allucinazioni”¹⁴¹, oltre a introdurre o perpetuare pregiudizi e ad esacerbare ulteriormente le disuguaglianze all'interno del sistema giudiziario.

Gli studiosi hanno evidenziato almeno sette categorie di rischi che possono derivare dall'uso della GenAI, tra cui la compromissione dell'uguaglianza¹⁴², la violazione dei diritti di proprietà intellettuale, la fuga di informazioni, l'uso dannoso, le minacce alla sicurezza, l'illusione del modello e i rischi ambientali/sociali e normativi. Nonostante gli sforzi in corso, nessuna categorizzazione unificata dei rischi dell'IA copre in modo esaustivo l'intero spettro di rischi, tenendo conto delle prospettive dell'industria e del governo¹⁴³.

Nello specifico ambito legale, le applicazioni di GenAI richiedono attenzione, soprattutto per i seguenti quattro rischi significativi: rischio di sicurezza (sicurezza dei dati), rischio di riservatezza, rischio di uso non etico e di violazione della proprietà intellettuale, rischio di non controllabilità delle decisioni legali e rischio di frode attraverso la formazione con informazioni inaffidabili o non veritiere.

Tali rischi sono il frutto di diversi limiti che evidenziano le prestazioni degli LLM¹⁴⁴.

¹³⁹ Op. ult. cit., pp. 14 ss.

¹⁴⁰ S. ZIPOLI CAIANI, *A cosa pensano le macchine? Efficienza e opacità nelle reti neurali artificiali*, in *Intelligenza artificiale. Sfide etiche e teoriche*, a cura di M. GALLETI, S. ZIPOLI CAIANI, il Mulino, Bologna, 2024, 21–44.

¹⁴¹ S. WACHTER, e al., *Do large language models have a legal duty to tell the truth?*, in *Royal Society Open Science*, 8, 2024, 1–38, 1.

¹⁴² I problemi di discriminazione e pregiudizio sono ben noti da tempo, come illustrato da J. ANGIN, e al., *Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks*, ProPublica, 23–05–2016 (reperibile in: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>).

¹⁴³ Y. ZENG, e al., *AI Risk Categorization Decoded (AIR 2024): From Government Regulations to Corporate Policies*, in *arXiv preprint arXiv:2406.17864*, 2024, 32–40.

¹⁴⁴ Per un elenco dei limiti: 1) illusione e finzione (“allucinazione”); 2) discriminazione algoritmica e bias; 3) mancanza di affidabilità; 4) non interpretabilità; 5) mancanza di empatia; 6) complesso reperimento di risorse; 7) incapacità di guardare al futuro (cattura da parte dei dati di addestramento); 8) limiti di sicurezza e riservatezza (sicurezza dei dati). Come abbiamo già evidenziato, i LLM come ChatGPT sono inclini alle “allucinazioni”, un particolare errore per cui l'output generato (ad esempio un testo) appare coerente in superficie ma può essere errato o completamente inventato. Il termine “allucinazione” si riferisce alla generazione di informazioni false, insensate o imprecise da parte di grandi

La limitazione più significativa, tipica di GenAI, consiste nel rischio che alcuni risultati siano falsificati o fittizi. Ciò è dovuto in parte alla configurazione dell’algoritmo e al rischio di correlazioni errate tra i dati. Inoltre, la mancanza di trasparenza nella programmazione dell’algoritmo e nel collegamento dei dati contribuisce a un’oscurità generale e complica la convalida della risposta fornita.

Un tipico esempio di “illusione o finzione” prodotta da un LLM in ambito giuridico è il caso *Mata c. Avianca*, deciso nel 2023 dal Tribunale federale del Distretto Sud di New York¹⁴⁵. Il caso aveva origini ordinarie: Roberto Mata aveva avviato un’azione nei confronti della compagnia aerea Avianca per un infortunio subito durante un volo internazionale nel 2019. Nella fase di opposizione alla richiesta di rigetto avanzata da Avianca, l’avvocato Steven Schwartz — con oltre trent’anni di esperienza — aveva utilizzato ChatGPT per svolgere la ricerca giurisprudenziale, convinto di disporre di un “motore di ricerca evoluto”¹⁴⁶. Il risultato fu disastroso: la memoria depositata conteneva numerosi precedenti giurisprudenziali inesistenti, con citazioni interne e virgolettati interamente inventati¹⁴⁷.

L’altro grande problema che hanno questi strumenti riguarda i possibili esiti discriminatori delle loro inferenze. Come ogni “famiglia” di algoritmi basati sull’apprendimento automatico, gli LLM possono perpetuare problemi, errori e pregiudizi già presenti nei database di addestramento. I dati utilizzati dagli algoritmi per l’addestramento e la messa a punto, così come la costruzione del meccanismo stesso, potrebbero essere basati su pregiudizi o preconcetti¹⁴⁸. Un problema che può condizionare l’impostazione dell’algoritmo, ad esempio attraverso l’inclusione o l’esclusione di elementi che identificano o fanno riferimento a una categoria protetta a rischio di discriminazione, senza possibilità di appello.

modelli linguistici (LLM) o altri sistemi di intelligenza artificiale generativa. Per saperne di più, vedere N. MALEKI, e al., *AI Hallucinations: A Misnomer Worth Clarifying*, in *arXiv preprint arXiv:2401.06796*, 2024, pp. 1 ss.

¹⁴⁵ *Mata v. Avianca*, 2023 WL 4114965 (S.D.N.Y. June 22, 2023) <https://www.casemine.com/judgement/us/6499235932898b72d5cfb1aa>

¹⁴⁶ Il problema emerse quando gli avvocati di Avianca notificarono al tribunale di non essere riusciti a individuare alcuni dei casi citati. Il giudice Kevin Castel ordinò alla difesa di produrre copia delle sentenze citate: Schwartz fornì copie di documenti che ChatGPT gli aveva assicurato essere reali e reperibili in banche dati come LexisNexis e Westlaw. Si trattava, anche quelle, di documenti falsi. Schwartz dichiarò in udienza di aver agito nella “falsa convinzione che ChatGPT non potesse fabbricare casi di sana pianta”, ritenendo piuttosto che si trattasse di sentenze non pubblicate o difficili da reperire. La condotta successiva aggravò ulteriormente la situazione. Solo quando fu avviato il procedimento per sanzioni, il legale ammise le proprie responsabilità; il giudice rilevò che il quadro sarebbe stato ben diverso se la rivelazione fosse avvenuta tempestivamente. Il tribunale ravvisò la malafede del legale in una serie di atti di “consapevole omissione e dichiarazioni false e fuorvianti rese alla Corte”. C.A. NORTON, N.B. RAPOPORT, *Doubling Down on Dumb: Lessons from Mata v. Avianca Inc.*, in *American Bankruptcy Institute Journal*, 8, 2023, 24–61; P.B. LADKIN, *Involving LLMs in legal processes is risky: An invited paper*, in *Digital Evidence & Elec. Signature L. Rev.*, 1, 2023, 40–46.

¹⁴⁷ Le conseguenze furono significative sotto diversi profili. Il giudice Castel dispose il rigetto della domanda del cliente e inflisse agli avvocati e allo studio legale una sanzione pecuniaria di 5.000 dollari. Fu inoltre imposto loro di inviare al loro assistito e ai giudici i cui nomi erano stati associati alle sentenze false una copia dell’ordinanza, la trascrizione dell’udienza e gli atti che avevano generato l’errore. Il caso *Mata c. Avianca* è oggi considerato il *leading case* sulle conseguenze dell’uso improprio dell’intelligenza artificiale generativa negli atti processuali, e ha innescato un ampio dibattito etico e deontologico nell’intera comunità giuridica statunitense. Vale la pena sottolineare che il tribunale non contestò, in sé, l’utilizzo di strumenti di intelligenza artificiale, riconoscendo che nulla vieta in astratto il ricorso a tecnologie affidabili come ausilio alla ricerca giuridica. Il nucleo della violazione era altro: l’omissione del dovere di verifica, che le norme processuali — in particolare la Rule 11 delle Federal Rules of Civil Procedure — impongono a ogni avvocato a garanzia dell’accuratezza degli atti depositati.

¹⁴⁸ È stato dimostrato che anche la GenAI, in alcuni casi, risente delle impostazioni culturali e dei pregiudizi dei suoi creatori e della limitazione dei dataset. S. WACHTER, B. MITTELSTADT, C. RUSSELL, *Do large language models have a legal duty to tell the truth?*, cit.

Questa limitazione, di cui abbiamo già parlato con riguardo al caso *Loomis*, richiama alla mente l'idea weberiana di “calcolabilità giuridica”¹⁴⁹, che è percepita come una priorità auspicabile, soprattutto in un contesto storico che sembra dominato e travolto da un cambiamento giurisprudenziale instabile e perenne. D'altra parte, l'uso delle macchine comporta il rischio di un sistema che non solo non è in grado di creare nuovi argomenti giuridici, ma si riduce alla ripetizione stanca e stagnante di sé stesso, perdendo la capacità di adattamento alla realtà, con il rischio di svalutare le peculiarità dei singoli casi e di standardizzare le formule per la gestione delle questioni giuridiche.

Grazie alla loro capacità di apprendere da quantità impressionanti di dati, gli LLM sono in grado di generare risultati sofisticati. Tuttavia, sebbene possa sembrare che il sistema sia in grado di ragionare in alcuni aspetti, non comprende come il testo che riproduce si colleghi alla realtà materiale, poiché non ha osservato un modello del mondo reale. Questo è un punto cruciale. Le macchine ragionano per inferenza e non per deduzione causale e, poiché non sono umane, mancano di empatia e comprensione emotiva¹⁵⁰. Bender e Koller lo illustrano con ciò che chiamano “The Octopus Test”¹⁵¹. In questo senso, gli LLM rivelano una questione critica relativa all'autorità e alla legittimità giuridica. In generale, la complessità del diritto non può essere ridotta in modo computazionale¹⁵². Le norme giuridiche non possono essere incapsulate in termini matematici e la loro interpretazione è intrinsecamente soggetta ad ambiguità.

L'usuale accostamento tra algoritmi e ragionamento giuridico si riflette nell'idea che entrambi impieghino un approccio basato su sequenze logiche. Di conseguenza, la convinzione che la logica giudiziaria possa essere allineata alla metodologia delle tecnologie digitali dovrebbe costituire un quadro ideale per strutturare il discorso forense. Tuttavia, gli algoritmi applicano la precisione dei loro automatismi a un mondo umano imperfetto, dove il calcolo incontra l'incertezza e l'imprevedibilità insite nel comportamento umano.

D'altra parte, l'oggetto del ragionamento giuridico non si esaurisce nella ripetizione seriale e tipica di fatti che, per quanto suscettibili di quantificazione, restano pur sempre caratterizzati dalla peculiarità di essere aderenti a un caso concreto. Da qui, nel prendere decisioni, emerge la necessità di farsi guidare dal criterio qualitativo espresso dalla meritevolezza dell'interesse piuttosto che da criteri quantitativi che si rivelano insufficienti.

Inoltre, e non secondariamente, occorre ricordare che il diritto è un sistema profondamente sociale aperto verso il futuro: la conoscenza giuridica — fatti, norme e

¹⁴⁹ C.V. GIABARDO, *The Judge and the Algorithm (in Defense of the Human Nature of Judging)*, 2021, (reperibile in: <https://ssrn.com/abstract=3859922>); N. IRTI, *Per un dialogo sulla calcolabilità giuridica*, in *Calcolabilità giuridica*, a cura di A. CARLEO, il Mulino, Bologna, 2017, 17–28; G. ALPA, *Note sulla calcolabilità nel diritto nordamericano*, in *Calcolabilità giuridica*, a cura di A. CARLEO, il Mulino, Bologna, 2017, 17–28; M. DE FELICE, *Calcolabilità e probabilità. Per discutere di 'incontrollabile soggettivismo della decisione'*, in *Calcolabilità giuridica*, a cura di A. CARLEO, il Mulino, Bologna, 2017, 37–55;

¹⁵⁰ C'è un ampio consenso tra gli informatici e persino tra i giuristi sul fatto che le macchine hanno poteri di calcolo e di accumulo di dati incomparabili con gli esseri umani. Tuttavia, non ragionano come gli esseri umani: accumulano ed elaborano dati di cui non comprendono il significato. Questo vale anche per software come ChatGPT e il resto delle cosiddette IA generative: imitano i testi umani senza cogliere il senso di ciò che fanno. E.J. LARSON, *The Myth of Artificial Intelligence: Why computers can't think the way we do*, Harvard University Press, Cambridge, MA, 2021; G. GIGERENZER, *How to Stay Smart in a Smart World: Why Human Intelligence Still Beats Algorithms*, The MIT Press, Cambridge, MA, 2022.

¹⁵¹ E.M. BENDER, A. KOLLER, *Climbing towards NLU: On meaning, form, and understanding in the age of data*, in 2020 (reperibile in: <https://aclanthology.org/2020.acl-main.463.pdf>). Il testo descrive un esperimento mentale: due naufraghi (A e B) comunicano via cavo sottomarino, e un polpo superintelligente (O) intercetta i messaggi. Pur non conoscendo l'inglese, O impara a prevedere le risposte di B attraverso il riconoscimento di pattern statistici. La conclusione è che saper riprodurre risposte credibili non equivale a comprendere il significato — una distinzione rilevante nel dibattito sull'intelligenza artificiale.

¹⁵² A. GARAPON, J. LASSÈGUE, *Justice digitale: Révolution Graphique et Rupture Anthropologique*, 1re édition., Presses universitaires de France, Paris, 2018.

conoscenza pratica — viene costruita dagli attori che operano al suo interno, in un contesto quotidiano fatto di ruoli, scadenze, relazioni e dinamiche interpersonali. Questo ambiente complesso influenza inevitabilmente come i casi vengono impostati e, di conseguenza, come vengono prese le decisioni¹⁵³. Si potrebbe pensare che contesti più formali, come il controllo giurisdizionale o l'appello — dove si discute di diritto più che di fatti — siano invece governati da regole chiare, e dunque più compatibili con sistemi algoritmici o di intelligenza artificiale. Ciò non è vero: anche in questi ambiti, fattori sociali determinano quali tra i molti precedenti disponibili diventino effettivamente rilevanti e influenti.

La socialità del diritto significa anche “apertura”. La legge deve poter essere messa in discussione ed evolversi. Al contrario, l'IA è un sistema chiuso. Le idee, i valori e i principi sono integrati in un sistema di IA sigillato ermeneuticamente che può elaborare solo i dati già inseriti¹⁵⁴. Ciò è particolarmente vero in relazione ai sistemi basati su LLM che riproducono solo ciò che è già presente nel set di dati, senza alcun riferimento a realtà materiale. Non può esserci sviluppo del diritto o cambiamento creativo, per non parlare delle sentenze che, con interpretazioni fuori dall'ordinario, hanno dato il via a cambiamenti sociali importantissimi in tanti ambiti (la tutela dei nuovi diritti, l'aborto, il divorzio, i diritti delle minoranze, degli esclusi e così via).

Al contrario, come attenta dottrina evidenzia, “gli LLM non sembrano fornire prova di conoscere la natura del diritto”, considerandolo, forse, in modo più semplicistico, come un tradizionale “comando sostenuto da sanzioni in stile Austin”, in cui regole chiare possono essere applicate a “fatti” semplici, ovvero come un semplice problema statistico¹⁵⁵. Un tale approccio implica negare, nascondere o distorcere tutte le relazioni e le interazioni sociali sottostanti che, come molti teorici hanno sostenuto da diverse prospettive, sono coinvolte nei sistemi giudiziari, propriamente detti.

In questi termini, riportandoci al punto di partenza relativo ai problemi tecnici derivanti dall'uso degli LLM nell'ambito giudiziario, si aprono nuovi interrogativi sull'accessibilità degli algoritmi al contraddittorio, sulle indagini preliminari, sulla formazione delle prove, sull'indipendenza del giudice e sulla sua capacità di tenere conto dei precedenti, senza mai perdere di vista le novità della realtà. Non si tratta più di appassionate discussioni teoriche, ma di prospettive scientifiche volte a sondare la praticabilità di un sistema giudiziario governato con l'ausilio di IA progressivamente perfezionate.

L'inserimento di informazioni giudiziarie per sollecitare la macchina non è un'operazione neutra. Una volta trasmesse al fornitore del sistema, le informazioni possono essere utilizzate come dati di addestramento per gli utenti futuri e per generare risultati futuri. Ciò potrebbe comportare una violazione dei dati personali utilizzati nel prompt o, peggio, la divulgazione involontaria di informazioni classificate o comunque sensibili (cybersecurity). I server o i cloud dei fornitori di GenAI sono sconosciuti e generalmente collocati al di fuori del territorio dell'UE.

La necessità di trovare un equilibrio tra l'esigenza di proteggere e incentivare lo sviluppo tecnologico e quella di rispettare i diritti degli individui assume oggi un'importanza fondamentale. È proprio qui, infatti, che si può cogliere l'importanza di sviluppare condizioni favorevoli al rispetto dei diritti dell'individuo attraverso un'informazione accurata, la trasparenza e una supervisione indipendente ed efficace delle tecnologie, senza creare ostacoli all'individuazione delle responsabilità e dei rimedi in caso di violazione dei

¹⁵³ J. MORISON, T. MCINERNEY, *When should a computer decide?: Judicial decision-making in the age of automation, algorithms and generative artificial intelligence*, in *Research handbook on judging and the judiciary*, a cura di S. TURENNE, M. MOUSSA, Edward Elgar Publishing, London, 2025, 54–87.

¹⁵⁴ J. MORISON, A. HARKENS, *Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making*, cit.

¹⁵⁵ Le citazioni sono tratte da J. MORISON, T. MCINERNEY, *When should a computer decide?: Judicial decision-making in the age of automation, algorithms and generative artificial intelligence*, cit., 83.

diritti. Altrettanto necessario è un processo che rispetti i principi di non discriminazione, uguaglianza e protezione dei dati personali.

Questi obiettivi evidenziano anche l'urgenza di regole giuridiche e principi etici chiamati a promuovere lo sviluppo di una tecnologia ormai indispensabile per la crescita economica e il benessere sociale, da un lato, e a garantire la tutela dei diritti fondamentali e dei principi di libertà e democrazia su cui si fonda lo Stato di diritto, dall'altro. Pertanto, la struttura intima della legge, insieme alle limitazioni specifiche associate alle LLM, scoraggia fortemente l'uso non autorizzato e amatoriale di tali strumenti.

3.3. Le strategie per risolvere tali problemi

Nel paragrafo precedente abbiamo visto i problemi e i limiti degli LLM. Il loro operare su basi statistiche li porta a riprodurre un "discorso" senza una vera comprensione del significato e delle categorie che esso rappresenta.

Esaminiamo ora le strategie emergenti e più consolidate per colmare il divario tra i risultati probabilistici dell'IA e le esigenze del ragionamento giuridico. Attraverso l'analisi della letteratura informatica¹⁵⁶, discuteremo di alcuni metodi ritenuti utili per affrontare alcune delle sfide emerse in precedenza. L'idea che emerge da questa letteratura è che, prima di applicare gli LLM agli ambiti giuridici, occorre disporre di una via di raccolta, di predisposizione e di accesso "affidabile" alle conoscenze giuridiche che l'IA deve elaborare. Ciò richiede tecniche che adattino i parametri interni del modello alle sfumature giuridiche o che fondino le sue risposte su fonti esterne autorevoli, migliorando l'accuratezza fattuale e riducendo le allucinazioni.

La dottrina organizza le strategie per affrontare le sfide legate all'applicazione degli LLM al diritto in sei grandi categorie. Insieme, queste sei categorie formano un quadro complementare e sinergico, ciascuna delle quali mira a modalità di errore distinte, sostenendo collettivamente l'integrazione responsabile dell'IA nella pratica legale.

La prima raggruppa i "metodi di ragionamento strutturato e di deliberazione", tra cui la scomposizione dei compiti, i prompt concatenati e l'albero dei pensieri. Tali metodi guidano il modello oltre la semplice generazione di testo verso un'analisi giuridica più sistematica e graduale, producendo un percorso logico esplicito e verificabile.

La seconda è composta dalle "architetture ibride e agentiche", come i sistemi neurosimbolici e le pipeline multi-agente, che estendono ulteriormente queste capacità combinando la comprensione del linguaggio dell'LLM con motori deterministici basati su regole o distribuendo compiti complessi tra agenti specializzati che collaborano tra loro.

La terza categoria comprende le "tecniche di verifica e perfezionamento dei risultati", ovvero l'autovalutazione strutturata e il punteggio di affidabilità basato sul *logit*, che fungono quindi da controlli di qualità automatizzati, segnalando le incongruenze e quantificando l'incertezza prima che una risposta venga accettata. Questo metodo è molto utile per individuare le fonti del diritto e per capire se l'LLM dispone di sufficienti informazioni per rispondere al prompt.

La quarta categoria è tra le più importanti, perché raggruppa i "meccanismi di equità procedurale e di trasparenza". Questi comprendono *audit trail*, obblighi di divulgazione e supervisione umana obbligatoria, e garantiscono che qualsiasi coinvolgimento dell'IA nel processo decisionale rimanga visibile, contestabile e, in ultima analisi, subordinato al giudizio umano.

Nella quinta categoria sono raggruppate le "metriche di valutazione specifiche del settore", quali il richiamo delle fonti, l'accuratezza nell'identificazione della *ratio decidendi* e il peso dei diversi riferimenti (dottrinali, giurisprudenziali, ecc.). Tali metodi forniscono i

¹⁵⁶ Ci basiamo soprattutto sull'indagine da ultimo svolta da E. LINNA, T. LINNA, *Challenges for generative AI in legal reasoning*, in *Discover Artificial Intelligence*, 2026, 2731-0809.

rigorosi parametri di riferimento necessari per misurare e convalidare le prestazioni del sistema rispetto agli standard autentici del ragionamento giuridico.

L'ultima categoria è composta dalle “tecniche di adattamento al dominio e di fondamento fattuale”, che ancorano i risultati del modello a conoscenze verificate e specifiche del dominio, riducendo il rischio di allucinazioni ed errori giurisdizionali. In questa categoria rientrano il *fine tuning*, il RAG e il TAAR. Vediamole nello specifico, dato che si tratta degli strumenti oggi più noti e impiegati per fornire “affidabilità” e “reliance” agli LLM.

Il “fine tuning” non è, in realtà, una tecnica che aggiunge qualcosa al modello. Il *fine tuning* continua l'addestramento di un modello pre-addestrato per adattarlo al vocabolario, allo stile e alle sfumature specifiche del dominio giuridico. Consente a un LLM di ottenere una base più solida nella conoscenza giuridica ed è utile per cogliere la struttura e il linguaggio associati all'*adjudication* delle clausole generali. Tuttavia, la messa a punto da sola non può soddisfare tutti i requisiti a causa della natura variabile e aperta del contesto giuridico. Richiede inoltre un set di dati di alta qualità e comporta il rischio di “catastrofica dimenticanza”, che degrada le prestazioni in capacità al di fuori dell'ambito della messa a punto.

Molto più promettente è il “Retrieval-augmented generation” (RAG). Il RAG tenta di colmare la lacuna statistica dei modelli e anziché generare risposte solo a partire da parametri preaddestrati, abbinava un modulo di recupero — che estrae leggi, giurisprudenza o dottrina — a un modulo generativo che condiziona l'output su quei materiali. Il risultato è una sorta di ermeneutica computazionale: il significato emerge dall'ancoraggio nei testi autorevoli, non *ex nihilo*¹⁵⁷.

L'altro metodo è il “temporal & authority-aware retrieval” (TAAR), che consente, attraverso dei grafi, di adattare il RAG al dominio giuridico attraverso strategie sensibili al tempo e consapevoli dell'autorità. I grafi di conoscenza temporale indicizzano leggi e casi in base alle date di entrata in vigore, garantendo che l'LLM faccia riferimento alla versione corretta di una legge ed evitando allucinazioni temporali¹⁵⁸.

Insieme, queste strategie formano un quadro multilivello che affronta le sfide fondamentali di affidabilità, trasparenza ed equità legate all'implementazione dei modelli di linguaggio grande (LLM) in contesti legali, sebbene nessun approccio sia sufficiente da solo e la loro combinazione efficace rimanga sia un'area attiva di ricerca e sviluppo sia un'area di implementazione molto importante.

4. La disciplina europea e italiana sull'IA nel settore giustizia

La necessità di una guida sull'uso dell'IA nel settore della giustizia è accentuata dal nuovo regolamento (UE) sull'intelligenza artificiale n. 2024/1689, noto anche come “AI Act” del luglio 2024. L'IA Act introduce principalmente un “quadro di sicurezza dei prodotti” articolato in quattro categorie di rischio: inaccettabile, elevato, limitato e minimo.¹⁵⁹ I sistemi vietati, etichettati come inaccettabili per il rischio che comportano, sono elencati all'articolo 5¹⁶⁰.

¹⁵⁷ Il RAG è essenziale per i compiti che richiedono l'identificazione corretta del contesto giuridico o la citazione precisa delle fonti. La sua efficacia, tuttavia, dipende dalla qualità sia del sistema di recupero sia della base di conoscenza. I metodi RAG comuni si basano sulla somiglianza semantica anziché su quella giuridica, il che può comportare la trascuratezza di documenti rilevanti.

¹⁵⁸ La classificazione consapevole dell'autorità codifica le gerarchie giuridiche e le reti di citazioni nel processo di recupero, ad esempio ponderando le decisioni dei tribunali in base all'“autorità” dei precedenti, allineando il recupero alla dottrina dello *stare decisis* e alla gerarchia delle norme.

¹⁵⁹ L'AI Act considera solo alcuni aspetti dei possibili usi dell'IA. Cfr. B. PRAINSACK, N. FORGÓ, *New AI regulation in the EU seeks to reduce risk without assessing public benefit*, in *Nature Medicine*, 5, 2024, 1235–1237; C. NOVELLI, e

L'IA Act classifica i sistemi di IA ad alto rischio in due gruppi distinti. La prima categoria comprende i sistemi che fungono da componenti di sicurezza all'interno di prodotti già sottoposti a valutazioni di conformità ex-ante, come previsto dalla legislazione di armonizzazione dell'UE in materia di salute e sicurezza descritta nell'Allegato I.¹⁶¹ La seconda categoria è costituita da sistemi di IA autonomi ad alto rischio, che destano preoccupazione soprattutto per il loro potenziale impatto negativo sui diritti fondamentali.¹⁶² Per quanto riguarda l'attenzione ai diritti fondamentali, la l'IA Act, all'articolo 27, prevede una valutazione d'impatto sui diritti fondamentali (FRIA) per alcuni sistemi di IA ad alto rischio.

4.1. La “giustizia” nell'AI Act

Secondo le nuove norme europee, la “giustizia” è uno dei settori in cui l'apporto dell'IA si farà particolarmente sentire, sia fornendo vantaggi competitivi fondamentali alle imprese sia sostenendo risultati utili dal punto di vista sociale e ambientale¹⁶³.

al., *AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act*, in *Digital Society*, 1, 2024, 1–29; G. DE MINICO, *Giustizia e intelligenza artificiale: un equilibrio mutevole*, in *Rivista AIC*, 2, 2024, 85–108; G. MAZZINI, S. SCALZO, *The Proposal for the Artificial Intelligence Act: Considerations around Some Key Concepts*, in *La via europea per l'Intelligenza Artificiale*, a cura di C. CAMARDI, Wolters Kluwer, Milano, 2022, 21–52.

¹⁶⁰ Queste comprendono l'impiego dell'intelligenza artificiale per manipolare il comportamento umano in modi che minano l'autonomia individuale e sfruttano le vulnerabilità delle persone fisiche, in particolare considerando le persone che, per età, disabilità o status socioeconomico, possono essere vulnerabili. Inoltre, comprende l'utilizzo di sistemi di IA per la valutazione del rischio criminale, la previsione del comportamento degli individui senza la supervisione umana e la valutazione o la classificazione degli individui in base al loro comportamento sociale o alle loro caratteristiche personali (social scoring). Tuttavia, tali pratiche vengono affrontate in modo specifico quando comportano un trattamento avverso o sfavorevole in contesti sociali non correlati al contesto originale in cui i dati sono stati raccolti o quando tale trattamento è ingiustificato o sproporzionato. L'impiego della tecnologia di riconoscimento delle emozioni sul posto di lavoro e nelle istituzioni scolastiche è espressamente vietato, consentendo così la sua applicazione in altri contesti nonostante le preoccupazioni etiche intrinseche che solleva. Allo stesso modo, è vietata l'estrazione indiscriminata di immagini facciali da Internet o dalle telecamere a circuito chiuso per costruire database di riconoscimento facciale. Inoltre, il quadro giuridico non consente di classificare i dati biometrici in base ad attributi sensibili, tra cui le convinzioni politiche, religiose o filosofiche, l'orientamento sessuale o la razza. L'AI Act consente alle forze dell'ordine di utilizzare la tecnologia di riconoscimento facciale in tempo reale negli spazi pubblici, ma solo a condizioni rigorose, come l'acquisizione di un'autorizzazione giudiziaria preventiva, e in alcune circostanze specifiche: condurre ricerche mirate per le vittime, prevenire particolari e imminenti minacce terroristiche, o localizzare o identificare individui che sono stati condannati o sono sospettati di aver commesso determinati reati gravi.

¹⁶¹ All'interno di questa categoria è possibile trovare un'ampia gamma di prodotti, tra cui dispositivi medici, giocattoli, attrezzature per l'aviazione, automobili e ascensori.

¹⁶² Questi sistemi sono elencati in modo esaustivo nell'Allegato III e abbracciano otto ambiti applicativi specifici. Quest'ultima categoria rappresenta un aspetto particolarmente significativo dell'AI Act, poiché i sistemi dell'Allegato I sono già soggetti a quadri normativi dedicati. La seconda categoria comprende i sistemi utilizzati per l'accesso al lavoro o quelli impiegati nei settori dell'istruzione, della giustizia e dei servizi pubblici, per i quali sono previsti requisiti di valutazione della conformità. I sistemi ad alto rischio devono rispettare i “requisiti essenziali” delineati negli articoli da 8 a 15 dell'AI Act, che riguardano il sistema di gestione del rischio, la qualità dei dati e le misure di governance dei dati, la documentazione tecnica e le capacità di registrazione (automatizzata), le disposizioni in materia di trasparenza, la supervisione umana, la sicurezza informatica e il sistema di gestione della qualità. I requisiti dell'AI Act sono differenziati a seconda del tipo di operatori (fornitori, implementatori, distributori, importatori o rappresentanti autorizzati).

¹⁶³ Si veda il testo del Considerando n. 4 “L'IA è una famiglia di tecnologie in rapida evoluzione che contribuisce a un'ampia gamma di vantaggi economici, ambientali e sociali nell'intero spettro delle industrie e delle attività sociali. Migliorando la previsione, ottimizzando le operazioni e l'allocazione delle risorse e personalizzando le soluzioni digitali disponibili per individui e organizzazioni, l'uso dell'IA può fornire vantaggi competitivi fondamentali alle imprese e sostenere risultati vantaggiosi dal punto di vista sociale e ambientale, ad esempio nei settori dell'assistenza sanitaria, dell'agricoltura, della sicurezza alimentare, dell'istruzione e della formazione, dei media, dello sport, della cultura, della gestione delle infrastrutture, dell'energia, dei trasporti e della logistica, dei servizi pubblici, della sicurezza, della giustizia, dell'efficienza delle risorse e dell'energia, del monitoraggio ambientale, della conservazione e del ripristino della biodiversità e degli ecosistemi e della mitigazione e dell'adattamento ai cambiamenti climatici”.

Il considerando n. 61 afferma che alcuni “sistemi di IA destinati all’amministrazione della giustizia e dei processi democratici dovrebbero essere classificati come ad alto rischio, visto il loro impatto potenzialmente significativo sulla democrazia, sullo Stato di diritto, sulle libertà individuali e sul diritto a un ricorso effettivo e a un processo equo”. In particolare, il Considerando afferma che “per affrontare i rischi di potenziali pregiudizi, errori e opacità”, i “sistemi di IA destinati a essere utilizzati da un’autorità giudiziaria o per suo conto per assistere le autorità giudiziarie nella ricerca e nell’interpretazione dei fatti e del diritto e nell’applicazione del diritto a un insieme concreto di fatti” devono essere qualificati come “ad alto rischio”. Allo stesso tempo, anche i sistemi di IA “utilizzati da organismi alternativi di risoluzione delle controversie” devono essere considerati ad alto rischio “quando i risultati dei procedimenti alternativi di risoluzione delle controversie producono effetti giuridici per le parti”. Si sottolinea inoltre che “gli strumenti di IA possono sostenere il potere decisionale dei giudici o l’indipendenza giudiziaria, ma non dovrebbero sostituirlo” perchè in ogni caso, “il processo decisionale finale deve rimanere un’attività guidata dall’uomo”.

Nell’ottica dell’AI Act, non tutti i sistemi di IA usati nel settore giustizia sono considerati ad alto rischio. Sono esclusi quelli destinati ad “attività amministrative puramente accessorie che non influiscono sull’effettiva amministrazione della giustizia nei singoli casi”. Il considerando fornisce esempi di esclusione, tra cui rientrano i sistemi di IA destinati all’anonimizzazione o pseudonimizzazione di decisioni, documenti o dati giudiziari, comunicazione tra il personale, compiti amministrativi”.

L’allegato III, punto n. 8, specifica inoltre che “i sistemi di IA destinati a essere utilizzati 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 del diritto a un insieme concreto di fatti, o a essere utilizzati in modo simile nella risoluzione alternativa delle controversie”¹⁶⁴.

Sarà importante comprendere correttamente le attività coperte dall’espressione “ricercare e interpretare i fatti e la legge e applicare la legge a un insieme concreto di fatti”¹⁶⁵. L’AI Act delinea uno spettro applicativo potenzialmente ampio per i sistemi di IA in ambito giudiziario, che comprende attività di ricerca e di interpretazione dei fatti e del diritto, fino alla loro applicazione al caso concreto. Tuttavia, tanto il considerando introduttivo quanto il punto 8 dell’Allegato III utilizzano il verbo “assistere”, segnalando con chiarezza che la funzione dell’IA è meramente strumentale e servente: lo strumento algoritmico non può sostituire il giudice, né diventare esso stesso titolare della funzione giurisdizionale. Malgrado l’AI Act qualifichi come “ad alto rischio” i sistemi di intelligenza artificiale che assistono i giudici nell’interpretazione del diritto, nella valutazione di fatti e prove, nella determinazione della responsabilità o nell’adozione di decisioni

¹⁶⁴ Ai sensi del combinato disposto dell’art. 6, paragrafo 2, e dell’Allegato III, punto 8, dell’AI Act, rientrano nella categoria dei sistemi ad alto rischio tutti i sistemi di IA impiegati nell’amministrazione della giustizia in funzione di supporto all’attività giudiziaria: a titolo esemplificativo, i sistemi destinati all’interpretazione delle norme, alla valutazione delle prove, alla redazione di minute di sentenze o alla determinazione delle sanzioni. Tale qualificazione non è tuttavia assoluta. Il fornitore può sottrarsi al regime più stringente previsto per i sistemi ad alto rischio qualora dimostri — attraverso la procedura di cui all’art. 6, paragrafo 3 — che il sistema, pur rientrando in una delle categorie dell’Allegato III, non genera un rischio significativo per i diritti fondamentali, la salute o la sicurezza delle persone. In tal caso, il sistema è esentato dagli obblighi previsti per i sistemi ad alto rischio e soggiace unicamente al regime ordinario. Si tratta, in altri termini, di una presunzione relativa di alto rischio.

¹⁶⁵ Per quanto riguarda l’applicazione di queste disposizioni, l’articolo 74, paragrafo 8, prevede che gli Stati membri si avvalgano delle autorità di controllo della protezione dei dati (ai sensi del regolamento (UE) 2016/679 o della direttiva (UE) 2016/680), come l’italiano “Garante per la protezione dei dati personali”, come “autorità di vigilanza del mercato ai fini del presente regolamento” quando sono considerati “sistemi di IA ad alto rischio elencati nell’allegato III, punto 1, del presente regolamento, nella misura in cui i sistemi sono utilizzati per finalità di contrasto, gestione delle frontiere e giustizia e democrazia” e “sistemi di IA ad alto rischio elencati nell’allegato III, punti 6, 7 e 8, del presente regolamento”.

giurisdizionali, non definisce con sufficiente precisione il confine tra le attività in cui l'ausilio algoritmico è ammissibile e quelle che rimangono riservate in via esclusiva alla cognizione del giudice¹⁶⁶.

A un esame approfondito, in questa disciplina si possono individuare tre ordini di problemi. In primo luogo, l'assenza di limiti al potere discrezionale dei giudici nazionali sulle condizioni in cui una decisione può fondarsi — o non fondarsi — sull'output di un sistema di IA. In secondo luogo, il mancato richiamo alle garanzie fondamentali del giusto processo sancite dalle costituzioni nazionali (art. 111 Cost.), dalle norme processuali e dalla CEDU (artt. 6 e 13): difesa, contraddittorio, obbligo di motivazione. In terzo luogo, e in via generale, l'assenza di una norma che disciplini esplicitamente il rapporto tra il giudizio umano e l'assistenza algoritmica, affermando la prevalenza della riserva di umanità nella decisione giurisdizionale¹⁶⁷.

Gli aspetti connessi alla giustizia sono coperti anche da altre norme del Regolamento. I Considerando nn. 170 e 171 impongono di prevedere mezzi di ricorso efficaci per le persone sui cui diritti e libertà incide negativamente l'uso dei sistemi di IA. Inoltre, le persone interessate hanno diritto di ottenere una spiegazione chiara qualora la decisione si basi principalmente sugli *output* di sistemi di IA ad alto rischio e incida significativamente sui diritti fondamentali.

L'articolo 86 riconosce alle persone interessate che sono sottoposte alla (maggior parte dei) sistemi di IA ad alto rischio elencati nell'allegato III un (ulteriore) "diritto a una spiegazione", che comprende il "diritto di ottenere dal responsabile dell'installazione spiegazioni chiare e significative sul ruolo del sistema di IA nella procedura decisionale e sui principali elementi della decisione presa". Questo diritto, tuttavia, si applica solo se la decisione "produce effetti giuridici o incide in modo analogo e significativo sulla persona in questione in un modo che essa ritiene abbia un impatto negativo sulla sua salute, sulla sua sicurezza o sui suoi diritti fondamentali", e il diritto nazionale o dell'Unione può prevedere eccezioni a questo diritto.

L'AI Act stabilisce obblighi di trasparenza per i sistemi a rischio limitato, come i chatbot (articolo 50)¹⁶⁸. Infine, i sistemi a rischio minimo possono essere coperti da codici di condotta volontari. L'AI Act incarica la Commissione europea, gli Stati membri e l'Ufficio per l'IA (un ente di vigilanza di cui parliamo nella prossima sezione) di incoraggiare e facilitare lo sviluppo di codici di condotta per promuovere l'applicazione volontaria dei requisiti per sistemi di IA a rischio elevato.

I modelli generativi di base sono, in teoria, al di fuori della classificazione basata sul rischio¹⁶⁹, ma sono regolamentati all'interno della categoria di "IA per scopi generali" (GPAI), anch'essa elencata nell'AI Act¹⁷⁰. L'articolo 53 impone ai fornitori di modelli GPAI l'obbligo di sviluppare e mantenere la documentazione tecnica, di garantire la conformità

¹⁶⁶ Come rileva R. D'ANGIOLELLA, *Giustizia e intelligenza artificiale tra l'AI ACT e il disegno di legge italiano. L'insostituibilità del giudice, il metodo e l'argomentazione giuridica*, Giustiziainsieme, 30 luglio 2025 (reperibile in: <https://www.giustiziainsieme.it/articolo/3581-giustizia-e-intelligenza-artificiale-tra-lai-act-e-il-disegno-di-legge-italiano-rosita-dangiollella>)

¹⁶⁷ G. GALLONE, *L'improcrastinabile esigenza di tracciare una via "italiana" per l'intelligenza artificiale nel procedimento amministrativo. Opportunità e legittimità di un intervento regolatorio nazionale a corredo dell'AI Act*, Giustiziainsieme, 26 giugno 2025 (reperibile in: www.giustiziainsieme.it/it/diritto-e-innovazione/3544-una-via-italiana-per-lintelligenza-artificiale-nel-procedimento-amministrativo-giovanni-gallone).

¹⁶⁸ E. LONGO, F. PAOLUCCI, *The Article 50 of the AI Act and the Transparency Obligations: The Model and its Limitations*, in *La disciplina dell'Intelligenza artificiale*, a cura di O. POLLICINO, et al., Giuffrè, Milano, 2024, 275–304.

¹⁶⁹ L'IA generativa non faceva parte della proposta originale; è stata aggiunta dal Consiglio e dal Parlamento durante il trilogico e l'approvazione al fine di imporre ulteriori requisiti di trasparenza ai fornitori di modelli di base generativi. P. HACKER, e al., *Regulating ChatGPT and other Large Generative AI Models*, in 2023 (reperibile in: <https://dx.doi.org/10.1145/3593013.3594067>).

¹⁷⁰ G. OLIVATO, *Paving the path towards general purpose AI systems regulation in the AI Act: an analysis of the Parliament's and Council's proposals*, in *Medialaws*, 3, 2023, 50–71.

alla legge sul diritto d'autore dell'UE e di diffondere sintesi “sufficientemente dettagliate” sui contenuti utilizzati per l'addestramento dei modelli prima della loro immissione sul mercato.¹⁷¹

Un sistema di IA basato su modelli linguistici di grandi dimensioni (LLM), se utilizzato nell'ambito dell'amministrazione della giustizia (Allegato III, punto 8), può essere classificato come ad alto rischio. Ciò si verifica quando il sistema è progettato per supportare la magistratura – ad esempio nella ricerca giuridica, nell'interpretazione dei fatti o nell'applicazione del diritto a casi concreti – e svolge un ruolo decisivo nel processo decisionale ai sensi dell'articolo 6, paragrafo 3.

Al contrario, se il sistema non deve essere sviluppato o utilizzato in modo da influenzare materialmente il processo decisionale, lo scopo previsto del sistema di IA rientra nell'esenzione di cui all'articolo 6, comma 3, dell'AI Act. Tuttavia, dato l'abituale eccessivo affidamento all'automazione, non è facile per il fornitore di sistemi di IA appellarsi all'esenzione. Il fornitore deve prevedere i potenziali ruoli dell'automazione nella progettazione del sistema. In questo caso, ogni uso prevedibile dovrebbe essere considerato un produttore di rischi e quindi mitigato mediante istruzioni d'uso complete ed efficaci, che dovrebbero anche essere progettate per prevenire qualsiasi influenza materiale che potrebbe aumentare il rischio di impatti negativi sui diritti fondamentali.¹⁷² In questi casi, si dovrebbe considerare anche il diritto a una spiegazione, come previsto dall'articolo 86 (vedi sopra).

Un ultimo aspetto su questo tema. Le norme dell'AI Act relative all'IA rischiosa entreranno in vigore il 2 agosto 2026. Tuttavia, è molto probabile che questa scadenza venga posticipata. La Commissione europea, preso atto dello scenario geopolitico della competitività tecnologica e dell'obiettiva complessità dell'intero quadro normativo digitale dell'UE, ha presentato il 19 novembre 2025 un articolato pacchetto di proposte legislative — il cosiddetto *Digital Package* — volto a una significativa semplificazione dell'apparato regolatorio. Il pacchetto si compone di tre iniziative distinte: il *Digital Omnibus* (nelle sue due componenti, COM(2025)837 e COM(2025)836), la *Data Union Strategy* e i portafogli digitali europei per le imprese¹⁷³. Con riferimento specifico all'AI Act, la Commissione

¹⁷¹ Gli obblighi per le GPAI includono “redigere e tenere aggiornata la documentazione tecnica del modello, compreso il processo di formazione e di test e i risultati della sua valutazione”, “redigere, tenere aggiornata e rendere disponibile l'informazione e la documentazione ai fornitori di sistemi di IA che intendono integrare il modello di IA per scopi generali nei loro sistemi di IA”, “redigere e rendere disponibile al pubblico una sintesi sufficientemente dettagliata dei contenuti utilizzati per la formazione del modello di IA per scopi generali”. Sebbene questi standard minimi si applichino a tutti i modelli, i modelli GPAI identificati come “rischio sistemico” a causa delle loro “capacità di impatto elevato” sono soggetti a obblighi aggiuntivi. Tali obblighi comprendono la conduzione di valutazioni del modello, l'esecuzione di test avversari, la valutazione e la mitigazione dei rischi sistemici, la segnalazione di incidenti gravi e la garanzia di un livello adeguato di sicurezza informatica. Tuttavia, è essenziale notare che i fornitori di modelli GPAI classificati come a rischio sistemico possono condurre le proprie verifiche e valutazioni piuttosto che affidarsi a verifiche esterne e indipendenti di terzi. Inoltre, non vi è alcun requisito per la concessione di licenze pubbliche ex-ante. Cfr. N.A. SMUHA, K. YEUNG, *The European Union's AI Act: beyond motherhood and apple pie?*, in *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*, a cura di N.A. SMUHA, CUP, Cambridge, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4874852, 2025, 228–258; M. EBERS, *Truly Risk-Based Regulation of Artificial Intelligence-How to Implement the EU's AI Act*, 2024, SSRN (reperibile in: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4870387).

¹⁷² Per una visione simile, si veda I. CARNAT, *Addressing the Risks of Generative Ai for the Judiciary: The Accountability Framework (S) Under the EU Ai Act*, 2024, SSRN 4887438 (reperibile in: <https://ssrn.com/abstract=4887438> or <http://dx.doi.org/10.2139/ssrn.4887438>)

¹⁷³ Il Digital Omnibus incide in modo rilevante su diversi atti normativi già vigenti — tra cui il GDPR, la Direttiva ePrivacy, NIS2, il Data Act e, appunto, l'AI Act — con l'obiettivo dichiarato di ridurre il carico amministrativo complessivo di almeno il 25% per le imprese in generale e del 35% per le PMI, in linea con le raccomandazioni del rapporto Draghi sulla competitività europea. Sul tema si v. no: S. BARONCELLI, *Dopo il Rapporto Draghi: semplificazione, pacchetti Omnibus e potere regolatorio dell'Unione*, in *Osservatorio sulle fonti (online)*, 3, 2025, 5–

propone, nell'ambito del *Digital Omnibus*, di adeguare le tempistiche di applicazione delle norme sull'alto rischio, collegandole alla disponibilità degli strumenti di supporto, come gli standard armonizzati¹⁷⁴.

Nel medesimo segno di una vocazione costituzionale per un efficiente bilanciamento fra innovazione tecnologica e presidio dei diritti fondamentali, il 20 novembre 2025 la Commissione ha presentato il *Digital Justice Package 2030*, con la strategia per la transizione digitale dei sistemi giudiziari europei, diretta a rafforzare l'impiego dell'IA nella giustizia e a fornire una formazione specifica ai magistrati e al personale giudiziario¹⁷⁵.

4.2. La legge n. 132/2025

Sebbene l'AI Act sia un regolamento europeo, le norme previste al suo interno non sono del tutto direttamente applicabili, richiedendo, come previsto dalle stesse norme sovranazionali, un'attività nazionale volta a dare seguito alle previsioni di armonizzazione. Il legislatore nazionale ha infatti adottato principi e regole per la *governance* delle innovazioni tecnologiche nei sistemi giudiziari.

La legge n. 132 del 23 settembre 2025, recante *Disposizioni e deleghe al Governo in materia di intelligenza artificiale*, si iscrive esplicitamente nel quadro normativo europeo: l'art. 1, comma 2, stabilisce che «le disposizioni della presente legge si interpretano e si applicano conformemente al regolamento (UE) 2024/1689», riaffermando, nei principi generali, la dimensione antropocentrica dell'uso della tecnologia, nel rispetto dei diritti fondamentali e delle libertà costituzionali (art. 3).

Il fulcro della disciplina sull'uso dell'IA in ambito giudiziario — estesa anche alla giustizia penale — è contenuto nell'art. 15, costruito attorno a un principio cardine: «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» (comma 1). Risultano così salvaguardati il nucleo centrale della funzione giurisdizionale e il principio di responsabilità umana ultima, mentre l'impiego dell'IA rimane confinato alle aree di supporto e assistenza al decisore¹⁷⁶.

La norma concentra poi nel Ministero della Giustizia tre tipologie di competenze:

- la disciplina degli impieghi dei sistemi di IA «per l'organizzazione dei servizi relativi alla giustizia, per la semplificazione del lavoro giudiziario e per le attività amministrative accessorie» (comma 2);

21; H. RUSCHEMEIER, *The Omnibus Package of the EU Commission*, 2025, disponibile in: <https://verfassungsblog.de/the-omnibus-package-of-the-eu-commission/>

¹⁷⁴ In concreto, la proposta prevede di posticipare le date già fissate — il 2 agosto 2026 per i sistemi ad alto rischio (Allegato III) e il 2 agosto 2027 per quelli integrati in prodotti regolamentati (Allegato I) — subordinando l'entrata in vigore delle norme alla definizione di standard armonizzati e strumenti di supporto alla compliance. L'applicazione delle regole sui sistemi ad alto rischio potrà slittare, in ogni caso non oltre il 2 dicembre 2027 per i sistemi dell'Allegato III e il 2 agosto 2028 per quelli rientranti nell'Allegato I. Il meccanismo introdotto è di tipo condizionale: il termine massimo è fissato in 16 mesi dalla data di una formale conferma da parte della Commissione europea, il che significa che l'entrata in applicazione potrà essere scaglionata e posticipata, garantendo un avvio coordinato con il completamento del quadro tecnico. Restano invece immutate le scadenze già decorse, ossia il divieto di pratiche di IA vietate (operativo dal 2 febbraio 2025) e gli obblighi per i modelli di IA per finalità generali (GPAI), applicabili dal 2 agosto 2025.

¹⁷⁵ European Commission, *DigitalJustice@2030: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*, COM(2025) 802 final, 2025 (reperibile in: https://commission.europa.eu/document/download/95918716-ce7d-401b-b6d5-e23effae5b36_en?filename=JUST_template_comingsoon_standard_4.pdf).

¹⁷⁶ F. DAL CANTO, S. ROVELLI, *Il modello costituzionale di ordinamento giudiziario e la sfida della digitalizzazione della giurisdizione: il ruolo del Ministero della Giustizia e del Consiglio superiore della magistratura prima e dopo la legge italiana sull'intelligenza artificiale*, in *Questione giustizia*, 2025, 1–14.

- il potere autorizzatorio per «la sperimentazione e l’impiego dei sistemi di intelligenza artificiale negli uffici giudiziari ordinari, sentite le Autorità nazionali di cui all’articolo 20 [AgID e ACN]» (comma 3);
- l’elaborazione di linee programmatiche e la promozione di specifiche attività didattiche nella «formazione digitale di base e avanzata» dei magistrati e del personale amministrativo (comma 4).

Sul piano delle deleghe, l’art. 24 attribuisce al Governo due mandati distinti. Il primo — previsto dall’art. 1, comma 1 — riguarda l’adozione di uno o più decreti legislativi per l’adeguamento della normativa nazionale al regolamento (UE) 2024/1689, secondo i principi e criteri direttivi fissati dal comma 2. Il secondo, enunciato al comma 3, affida al Governo il compito di adottare, entro dodici mesi dall’entrata in vigore della legge, uno o più decreti legislativi per «adeguare e specificare la disciplina dei casi di realizzazione e di impiego illeciti di sistemi di intelligenza artificiale», con riferimento, dunque, al versante sanzionatorio e penalistico della materia.

I principi e criteri direttivi della delega appaiono ancora eccessivamente generici rispetto all’obiettivo di una effettiva armonizzazione del sistema processuale penale interno. Mancano, in particolare, prescrizioni puntuali destinate ai decreti attuativi in ordine ai poteri, ai divieti, alle condizioni, ai limiti e alle procedure per l’acquisizione, l’utilizzo e la valutazione delle informazioni generate da sistemi di intelligenza artificiale — inclusa quella generativa — sia nella fase delle indagini preliminari che nel giudizio.

A tali lacune di contenuto si aggiungono ulteriori profili problematici di carattere strutturale. Sul piano delle risorse, la consueta clausola di invarianza finanziaria (art. 27) segnala una sostanziale indifferenza del legislatore rispetto alla sostenibilità economica delle innovazioni tecnologiche introdotte, pur riconosciute come necessarie. Sul piano organizzativo, la ripartizione delle competenze tra il Ministero della Giustizia, AgID e ACN risulta eccessivamente frammentata, con il rischio di generare sovrapposizioni e conflitti di attribuzioni in una materia che richiederebbe, al contrario, una governance unitaria e coerente.

Le criticità più rilevanti, tuttavia, riguardano tanto la compatibilità del potere autorizzatorio (provvisorio) rimesso al Ministro della Giustizia quanto l’assenza di qualsiasi coinvolgimento del *Consiglio Superiore della Magistratura* nel processo complessivo di governance della digitalizzazione della giurisdizione¹⁷⁷. L’intero impianto risulta modellato secondo uno schema di tipo ministeriale, nonostante la materia incida direttamente sull’esercizio concreto e responsabile della funzione giudiziaria. Tale scelta legislativa solleva fondati interrogativi in ordine alla tenuta del principio di indipendenza della magistratura, che rischia di essere vulnerato non già attraverso interventi diretti sullo status dei giudici, bensì per via indiretta, mediante il controllo sugli strumenti tecnologici di cui essi si avvalgono nell’esercizio della giurisdizione. Inoltre, risulta alquanto complesso coordinare la perimetrazione nazionale individuata all’art. 15, comma 2, della legge con l’allegato III, punto 8, dell’AI Act. La distinzione tra “organizzazione dei servizi relativi alla giustizia”, “semplificazione del lavoro giudiziario” e “attività amministrative accessorie” risulta, nella pratica, assai più sfumata di quanto suggerisca il testo normativo.

Come abbiamo visto, il Regolamento europeo, al considerando 61, esclude dalla classificazione ad alto rischio solo i sistemi destinati ad attività “puramente accessorie, che non incidono sull’effettiva amministrazione della giustizia nei singoli casi”. L’avverbio *puramente* e la locuzione *non incidono* assumono così un peso ermeneutico decisivo, imponendo una valutazione rigorosa dell’effettiva influenza dell’automazione

¹⁷⁷ C. DE ROBBIO, *L’intelligenza artificiale e il lavoro del giudice. I rischi per l’autonomia e l’indipendenza della magistratura*, Giustiziainsieme, 4 febbraio 2026 (reperibile in: <https://www.giustiziainsieme.it/articolo/3797-l-intelligenza-artificiale-e-il-lavoro-del-giudice>).

sull'amministrazione della giustizia. Attività come la distribuzione automatizzata degli affari tra magistrati o la definizione algoritmica delle priorità processuali si collocano in un'evidente "zona grigia": formalmente riconducibili all'organizzazione del lavoro, esse implicano necessariamente operazioni di qualificazione giuridica che difficilmente possono dirsi "puramente" accessorie¹⁷⁸.

Rimangono esclusi in via totale i sistemi di IA generalisti che non sono stati né autorizzati né certificati per l'uso giudiziario¹⁷⁹.

5. I documenti di indirizzo circa l'uso della GenAI nei processi

Negli ultimi anni, molti organismi e istituzioni a livello internazionale e nazionale hanno elaborato documenti di indirizzo sull'uso delle tecnologie nei processi¹⁸⁰.

I documenti di indirizzo più recenti, alcuni dei quali verranno esaminati qui, sono dedicati all'impiego della GenAI nei processi. Quasi tutti convergono nel delineare un paradigma regolatorio improntato alla prudenza, alla gradualità e alla funzionalizzazione dell'innovazione tecnologica ai fini propri della giurisdizione.

La GenAI non è configurata quale possibile sostituto dell'attività interpretativa e decisoria del giudice, bensì quale strumento ausiliario destinato a operare entro un perimetro rigorosamente delimitato, a supporto delle attività preparatorie, conoscitive e organizzative che precedono la decisione¹⁸¹. Ne risulta una concezione marcatamente "servente" della tecnologia, la cui legittimità dipende dal mantenimento della piena titolarità umana del giudizio e dalla preservazione dell'autonomia della funzione giurisdizionale.

In tale prospettiva, il tratto forse più rilevante di questi documenti è la costante riconduzione del tema dell'uso della GenAI all'interno della cornice assiologica dello "Stato di diritto"¹⁸², mettendo i principi di trasparenza, spiegabilità, tracciabilità, responsabilità, controllabilità e supervisione umana non come semplici requisiti di buona amministrazione tecnica, bensì come condizioni di conformità costituzionale e internazionale dell'innovazione. Questi elementi sono ancora più essenziali in un contesto

¹⁷⁸ Op. ult. cit.

¹⁷⁹ Rispetto a tali sistemi, si pone un problema non secondario: che cosa accade degli usi non autorizzati da parte dei magistrati? Questa fattispecie non è esplicitamente sanzionata dalla legge. Si deve presumere che, in merito, siano applicabili le ordinarie misure disciplinari del d.lgs. n. 109/2006 e le regole processuali ordinarie. Sul primo piano si pongono questioni di inosservanza di norme organizzative e informatiche, di violazione dei doveri d'ufficio con possibile danno alle parti, di negligenza inescusabile e di indebito affidamento all'IA. Sul piano processuale, l'atto potrebbe essere inesistente in ambito civile per difetto genetico, o affetto da nullità assoluta in ambito penale ex art. 179 c.p.p., con conseguente regressione del procedimento. Questi rilievi sono evidenziati anche da F. IANNONE, *Prime riflessioni sull'IA: la legge 132/2025 letta in cornice eurounitaria e sovranazionale*, Giustiziainsieme, 7 novembre 2025 (reperibile in: www.giustiziainsieme.it/it/diritto-e-innovazione/3680-prime-riflessioni-intelligenza-artificiale-fernanda-iannone).

¹⁸⁰ Su questo argomento sia consentito rimandare all'esame in E. LONGO, *Giustizia digitale e Costituzione: riflessioni sulla trasformazione tecnica della funzione giurisdizionale*, cit., 285.

¹⁸¹ CSM, *Raccomandazioni sull'uso dell'intelligenza artificiale nell'amministrazione della giustizia*, 8 ottobre 2025 (reperibile in: <https://www.csm.it/portale/documents/21768/87316/linee+guida+intelligenza+artificiale+%28delibera+8+ottobre+2025%29.pdf/e0ed018a-43e2-9dd7-2420-e3d0632d658c?t=1772619014845>).

¹⁸² Segretariato Generale della Giustizia Amministrativa. Servizio per l'Informatica, *Intelligenza artificiale e Giustizia amministrativa: strategie di impiego, metodologie e sicurezza*, 4 ottobre 2024 (reperibile in: <https://assets.innovazione.gov.it/1751549488-ia-per-ga-report-sito-ga-2024.pdf>).

in cui si tutelano il giusto processo, il diritto di difesa, l'uguaglianza delle armi, la non discriminazione e la fiducia pubblica nel sistema giudiziario¹⁸³.

Sotto un profilo diverso ma complementare, i documenti di indirizzo mostrano una significativa convergenza nell'individuazione degli impieghi ritenuti più compatibili con la struttura del processo.

L'attenzione si concentra infatti su usi circoscritti, strumentali e organizzativamente sostenibili: analisi e classificazione documentale, supporto alla ricerca, estrazione di informazioni, sintesi, traduzione, anonimizzazione, gestione dei fascicoli e assistenza nella standardizzazione di attività ripetitive.

In questa chiave, l'uso della GenAI è concepito come fattore di incremento dell'efficienza e della capacità di trattare la complessità documentale, senza che ciò comporti alcuna legittimazione di automatismi decisionali o di surrettizie deleghe cognitive in ordine alla valutazione dei fatti e all'applicazione del diritto al caso concreto. La distinzione tra ausilio tecnico e funzione giurisdizionale diviene, dunque, il criterio ordinatore fondamentale dell'intero impianto prescrittivo.

Parallelamente, tali documenti insistono con particolare forza sui rischi specifici che l'impiego della GenAI comporta nel contesto processuale¹⁸⁴.

In questa prospettiva, la supervisione umana non può ridursi a formula rituale, ma deve tradursi in una verifica effettiva, competente e documentabile dei risultati prodotti dal sistema; al contempo, la tutela della riservatezza e della sovranità del dato giudiziario emerge quale preconditione imprescindibile di qualsiasi percorso di adozione responsabile¹⁸⁵.

Ne consegue che i documenti qui richiamati non si limitano a formulare divieti o cautele, ma prefigurano un vero e proprio modello istituzionale di governo della GenAI nei processi, fondato su sperimentazione controllata, integrazione tra competenze giuridiche e tecniche, formazione degli operatori, monitoraggio continuo e progressivo adattamento delle regole d'uso.

La questione dell'intelligenza artificiale generativa, in altri termini, non viene trattata come problema settoriale di innovazione digitale, bensì come banco di prova della capacità dell'ordinamento di assorbire nuove tecnologie senza alterare i presupposti di legittimazione della giurisdizione¹⁸⁶.

5.1. Le raccomandazioni del Csm

L'8 ottobre 2025 il Consiglio Superiore della Magistratura ha adottato alcune raccomandazioni rivolte ai magistrati in merito agli usi ammessi dell'IA¹⁸⁷, in particolare per delineare cosa fare in attesa dell'applicazione completa dell'AI Act.

Le raccomandazioni risolvono pragmaticamente le incertezze interpretative descritte di sopra mediante un elenco operativo¹⁸⁸, dichiaratamente non esaustivo, di impieghi

¹⁸³ L'adozione di sistemi di IA in ambito giudiziario è pertanto subordinata alla definizione di assetti di governance trasparenti, verificabili e sindacabili, tali da impedire che l'opacità tecnica si traduca in opacità giuridica, come mettono bene in luce le UNESCO, *Guidelines for the Use of AI Systems in Courts and Tribunals*, cit.

¹⁸⁴ L'inaffidabilità fattuale degli output, i fenomeni allucinatori, la propagazione o la "neutralizzazione apparente" dei bias, l'opacità dei modelli, la possibilità di un uso improprio di dati riservati, nonché il pericolo di uno slittamento funzionale da compiti amministrativi o meramente preparatori verso ambiti sostanzialmente decisionali, sono descritti come fattori che impongono cautele rafforzate. Si v. Advisory Committee on Artificial Intelligence and the Courts, *Annual Report to the Chief Judge and Chief Administrative Judge of the State of New York*, Dec 2025 (reperibile in: <https://www.nycourts.gov/LegacyPDFS/publications/pdfs/AI%20Annual%20Report%202025.pdf>)

¹⁸⁵ CSM, *Raccomandazioni sull'uso dell'intelligenza artificiale nell'amministrazione della giustizia*, cit.

¹⁸⁶ Come si è messo in evidenza in E. LONGO, *Giustizia digitale e Costituzione: riflessioni sulla trasformazione tecnica della funzione giurisdizionale*, cit., 325 ss.

¹⁸⁷ CSM, *Raccomandazioni sull'uso dell'intelligenza artificiale nell'amministrazione della giustizia*, cit.

dell'intelligenza artificiale che, allo stato attuale della normativa e nelle more della piena operatività dell'AI Act, possono considerarsi compatibili con i principi costituzionali e con le prescrizioni eurounitarie, a condizione che vengano implementati in modalità tracciata e sicura, con costante revisione umana, e nell'ambito degli applicativi forniti all'interno del dominio giustizia.

Tale elenco riflette una logica di fondo ben definita che abbiamo analizzato *supra* nei particolari: l'intelligenza artificiale può legittimamente assistere il magistrato in tutte quelle attività che, pur essendo strumentali o accessorie rispetto alla funzione giurisdizionale in senso stretto, richiedono un impegno significativo di tempo e di risorse cognitive, a condizione che il contributo algoritmico non si sostituisca mai alla valutazione umana, ma si limiti a supportarla e facilitarla.

Tra le molteplici attività suscettibili di essere assistite da strumenti di intelligenza artificiale, quella relativa alle ricerche su banche dati giurisprudenziali merita una riflessione autonoma e approfondita, collocandosi su un crinale sottile tra l'ammissibilità e la problematicità giuridica.

L'utilizzo dell'intelligenza artificiale per tale finalità presenta una natura intrinsecamente bifronte. Da un lato, essa può validamente assistere il magistrato nella consultazione delle banche dati istituzionali e commerciali, nella costruzione di stringhe di ricerca complesse e nella classificazione tematica del materiale reperito, configurandosi, in tal caso, come supporto tecnico-organizzativo riconducibile ai "compiti procedurali limitati" di cui all'art. 6, par. 3, lett. a), del Regolamento (UE) 2024/1689. Dall'altro lato, qualora i sistemi siano progettati per selezionare automaticamente la giurisprudenza ritenuta "più rilevante", per suggerire orientamenti interpretativi prevalenti o per generare schemi motivazionali fondati su pattern ricorrenti, si configura un impiego che incide potenzialmente sull'attività valutativa e sull'indirizzo giuridico, fuoriuscendo dall'ambito applicativo della deroga di cui al citato art. 6, par. 3 dell'AI Act.

Si impone, conseguentemente, una vigilanza stringente su tre piani distinti: la natura e l'architettura dei sistemi utilizzati; la trasparenza degli algoritmi di selezione e classificazione; il ruolo attivo e critico del magistrato nel vaglio dei risultati.

In ogni caso, l'output prodotto deve essere oggetto di valutazione e verifica autonome da parte del magistrato, quale titolare esclusivo della funzione giurisdizionale: l'automazione della ricerca non può sostituire la sensibilità giuridica necessaria alla corretta contestualizzazione del precedente. Sul piano delle condizioni strutturali di utilizzo, è parimenti necessario che le banche dati giurisprudenziali messe a disposizione del magistrato garantiscano una base dati completa, non discriminatoria e costantemente aggiornata, e che prevedano adeguate forme di controllo e supervisione da parte della magistratura nelle fasi di selezione, classificazione e aggiornamento delle sentenze.

Con riferimento specifico alla fase transitoria, anteriore alla piena operatività del quadro regolamentare europeo, il Csm chiarisce che potranno essere utilizzati i sistemi

¹⁸⁸ Gli ambiti compatibili individuati includono: le ricerche dottrinali, intese come assistenza nella consultazione di banche dati e nella costruzione di stringhe di ricerca finalizzate all'individuazione di contributi pertinenti; la sintesi di provvedimenti ostensibili e di contributi dottrinali per la creazione di abstract destinati alla classificazione e all'archiviazione tematica; le attività organizzativo-gestionali, quali la redazione di report statistici sull'andamento degli uffici, l'analisi di conformità tra programmi di gestione e dati di registro, la comparazione automatizzata di documenti, la predisposizione di bozze di relazioni su incarichi direttivi e semidirettivi, nonché la gestione dei calendari d'udienza sulla base dei carichi di lavoro e delle scadenze; il supporto agli uffici per i cosiddetti "affari semplici", mediante la redazione di bozze standardizzate da adattare al caso specifico da parte del magistrato o dei suoi collaboratori; il supporto ad attività giurisdizionali di natura gestionale attraverso il controllo della documentazione prodotta in atti, ove opportunamente anonimizzata; nonché una serie di attività tecnico-amministrative, tra cui la generazione automatica di materiali illustrativi, la produzione di tabelle e grafici estratti dai registri, la revisione linguistica e stilistica di bozze, la catalogazione e l'archiviazione tematica dei quesiti ai consulenti tecnici d'ufficio, la predisposizione automatizzata di calendari d'udienza e la traduzione assistita di testi giuridici, da sottoporre in ogni caso a verifica umana.

espressamente autorizzati dal Ministero della Giustizia ai sensi dell'art. 15, comma 2, della legge 23 settembre 2025, n. 132; gli strumenti forniti all'interno del dominio giustizia, quali le funzionalità di "Copilot" rese disponibili tramite il pacchetto Office365 ministeriale; nonché i progetti sperimentali condotti in ambiente protetto e sotto la supervisione congiunta del Ministero e del Consiglio, purché assistiti da preventiva anonimizzazione e da piena tracciabilità dei dati.

5.2. Le altre linee guida a livello internazionale

Già da tempo, di fronte alla possibile diffusione di decisioni giudiziarie che fanno uso di strumenti algoritmici, soprattutto in ambito penale, le organismi di tutto il mondo, a livello statale ed internazionale, hanno adottato diversi documenti etici e normativi per mitigare i possibili effetti negativi dell'IA.

Il primo e più importante documento è certamente la "Carta etica europea per l'uso dell'intelligenza artificiale nei sistemi di giustizia penale e negli ambienti connessi", adottata il 3 e 4 dicembre 2018 dalla Commissione per l'efficienza della giustizia del Consiglio d'Europa (CEPEJ). La Carta etica è il primo importante documento a livello internazionale ad affrontare le questioni relative all'uso di strumenti di IA in ambito giudiziario¹⁸⁹. Per la prima volta, l'importanza dell'intelligenza artificiale nella giustizia, in termini di efficienza e qualità della gestione, viene riconosciuta sia per i suoi benefici sia per i suoi effetti negativi¹⁹⁰.

La Carta etica è precedente all'avvento della GenAI. Oggi la situazione è leggermente cambiata e le possibilità di utilizzare l'IA nei processi e, in generale, nelle professioni legali si sono moltiplicate, fino a essere alla portata di (quasi) tutti, come documentato sopra parlando dei chatbot di IA alimentati dagli LLM. Il documento, tuttavia, rimane un punto essenziale di documentazione degli usi ammissibili da parte dei giudici e degli altri professionisti del diritto.

Dopo la Carta etica, il Consiglio d'Europa si è pronunciato anche su tema della GenAI. Il Gruppo di lavoro CEPEJ su *Cybergiustizia e Intelligenza Artificiale* ha pubblicato anche una "nota informativa" sull'"Uso dell'Intelligenza Artificiale Generativa (IA) da parte dei professionisti giudiziari in un contesto lavorativo". Il documento fornisce linee guida e considerazioni per un uso appropriato degli strumenti basati sull'IA generativa: delinea le migliori pratiche per garantire un utilizzo responsabile ed efficace, evidenziando al contempo le situazioni in cui il loro uso può essere inappropriato o vietato.

Oltre al Consiglio d'Europa, numerose altre istituzioni stanno cercando di analizzare gli effetti della GenAI sulla giustizia, e in particolare sul processo giudiziario¹⁹¹.

Con l'"Opinion on Regulating and Strengthening the Judicial Application of Artificial Intelligence" (parere sulla regolamentazione e il rafforzamento dell'applicazione giudiziaria dell'intelligenza artificiale), pubblicato l'8 dicembre 2022, la Corte Suprema del Popolo cinese ha stabilito chiaramente i principi per i processi assistiti dall'IA e ha sottolineato che, a prescindere dal livello di sviluppo tecnologico, l'IA non dovrebbe

¹⁸⁹ La Carta fornisce un quadro di principi che possono guidare i politici, i legislatori e i professionisti della giustizia nell'affrontare il rapido sviluppo dell'IA nei processi giudiziari nazionali. La Carta ha identificato cinque principi fondamentali: rispetto dei diritti fondamentali; non discriminazione; qualità e sicurezza dei dati; trasparenza, imparzialità ed equità; "sotto il controllo dell'utente". European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe, *European Ethical Charter on the use of artificial intelligence in judicial systems*, 3-4/12/2018 (reperibile in: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>)

¹⁹⁰ S. QUATTROCOLO, *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, cit..

¹⁹¹ R. SCHWORER, *Ethical Implications of Utilizing Generative Artificial Intelligence in the Legal Profession: A Cautionary Note on Potential Model Rules of Professional Conduct Violations*, cit..

sostituire i giudici nel giudizio, garantendo così che i poteri giudiziari siano sempre esercitati dall'organizzazione giudiziaria e non dal processo decisionale dell'IA¹⁹².

Nel luglio del 2024, l'American Bar Association (Comitato permanente per l'etica e la responsabilità professionale) ha emesso un "parere formale" sugli strumenti di intelligenza artificiale generativa, volto a garantire la tutela dei clienti quando gli avvocati utilizzano la GenAI¹⁹³. Gli avvocati devono considerare "i loro obblighi etici applicabili, compresi i doveri di fornire una rappresentanza legale competente, di proteggere le informazioni sui clienti, di comunicare con i clienti, di supervisionare i loro dipendenti e agenti, di avanzare solo richieste e contestazioni meritevoli, di garantire la sincerità nei confronti del tribunale e di applicare onorari ragionevoli".

Questa iniziativa è solo l'ultima di una lunga lista. Nel gennaio 2024 l'Ordine degli Avvocati dello Stato di New York ha istituito una Task Force sull'intelligenza artificiale, che ha prodotto un lungo rapporto con uno studio approfondito e raccomandazioni per gli avvocati su come utilizzare questi strumenti¹⁹⁴. Il rapporto descrive l'impatto legale, sociale ed etico dell'intelligenza artificiale (AI) e dell'AI generativa sulla professione legale. Il Rapporto esamina anche i software basati sull'IA, la tecnologia dell'IA generativa e altri strumenti di apprendimento automatico che possono migliorare la professione, ma anche generare rischi imprevisti, come l'incapacità dei singoli avvocati di comprendere questa tecnologia sconosciuta e problemi per i tribunali, preoccupati per l'integrità del processo giudiziario. Inoltre, il Rapporto formula raccomandazioni, tra cui la proposta di linee guida per un uso responsabile dell'IA.

Anche l'ultimo rapporto annuale 2025 del Chief Administrator dello Stato di New York riporta una novità. Con la creazione dell'Advisory Committee on Artificial Intelligence and the Courts è stato avviato un esame specifico sull'impatto dell'intelligenza artificiale nei tribunali. Il documento analizza le opportunità e i rischi in ambiti quali l'accesso alla giustizia, la gestione amministrativa, i profili etici, i possibili bias, l'affidabilità delle prove e l'uso dell'AI nella redazione degli atti giudiziari. Il rapporto si propone quindi come una guida strategica e operativa per favorire un'integrazione responsabile, trasparente e sicura dell'AI nel sistema giudiziario.

Linee guida su come i giudici devono utilizzare la GenAI sono apparse in altre giurisdizioni, come Australia¹⁹⁵, Canada¹⁹⁶ e Nuova Zelanda¹⁹⁷.

Anche i principali organi giudiziari inglesi e gallesi hanno redatto una "guida" nel 2023¹⁹⁸. La guida mette in guardia i giudici e sottolinea con forza l'uso responsabile della

¹⁹² W. ZHENG, *Research on Generative Artificial Intelligence Legal Profession Substitution*, in *Mod. L. Rsch.*, 3, 2023, 32–40.

¹⁹³ *Formal Opinion 512 "Generative AI Tools"*, 29/7/2024 (reperibile in: https://www.americanbar.org/groups/professional_responsibility/committees_commissions/ethicsandprofessionalresponsibility/)

¹⁹⁴ NYSBA Task Force on Artificial Intelligence, *Report and Recommendations of the New York State Bar Association Task Force on Artificial Intelligence*, 4 Apr 2024 (reperibile in: <https://nysba.org/app/uploads/2022/03/2024-April-Report-and-Recommendations-of-the-Task-Force-on-Artificial-Intelligence.pdf>)

¹⁹⁵ Queensland Courts, *The use of Generative Artificial Intelligence (AI): Guidelines for responsible use by non-lawyers*, 13/5/2024 (reperibile in: <https://www.courts.qld.gov.au/going-to-court/using-generative-ai>); *Guidelines for litigants. Responsible use of artificial intelligence in litigation*, 5/2024 (reperibile in: <https://www.supremecourt.vic.gov.au/forms-fees-and-services/forms-templates-and-guidelines/guideline-responsible-use-of-ai-in-litigation>)

¹⁹⁶ Federal Court of Canada, *Notice to the Parties and the Profession. The Use of Artificial Intelligence in Court Proceedings*, 7/5/2024 (reperibile in: <https://www.fct-cf.gc.ca/Content/assets/pdf/base/FC-Updated-AI-Notice-EN.pdf>); Federal Court of Canada, *Interim Principles and Guidelines on the Court's Use of Artificial Intelligence*, 20/12/2023 (reperibile in: <https://www.fct-cf.gc.ca/en/pages/law-and-practice/artificial-intelligence>)

¹⁹⁷ Courts of New Zealand, *Guidelines for use of generative artificial intelligence in Courts and Tribunals (Judges, Judicial Officers, Tribunal Members and Judicial Support Staff)*, 7/12/2023 (reperibile in: <https://www.courtsofnz.govt.nz/assets/6-Going-to-Court/practice-directions/practice-guidelines/all-benches/20231207-GenAI-Guidelines-Judicial.pdf>).

tecnologia¹⁹⁹, in particolare considerando le questioni di accuratezza (dei documenti prodotti da avvocati e giudici)²⁰⁰, riservatezza, cybersecurity (segretezza delle informazioni e cybersecurity)²⁰¹ e protezione dei dati²⁰².

Il documento, redatto dai vertici della magistratura inglese e gallese, mette in guardia i titolari di cariche giudiziarie sul materiale prodotto a loro nome. L'assunto di partenza riguarda il fatto che "i giudici non sono generalmente tenuti a descrivere le ricerche o il lavoro preparatorio svolto per produrre una sentenza".

Un elemento importante della Guida riguarda il rapporto con le parti esterne, gli "utenti" dei tribunali, cioè i professionisti legali, i consulenti e le parti stesse²⁰³.

La guida rappresenta solo la fase iniziale di una serie di iniziative che verranno intraprese per assistere il sistema giudiziario nella gestione dell'IA. È stato sottolineato che queste iniziative saranno rivalutate con il progredire della tecnologia. Si sottolinea che "l'uso dell'IA da parte o per la magistratura deve essere in linea con il dovere generale della magistratura di sostenere l'integrità dell'amministrazione della giustizia".

¹⁹⁸ Courts and Tribunals Judiciary, *Artificial Intelligence (AI). Guidance for Judicial Office Holders*, 2023 (reperibile in: <https://www.judiciary.uk/guidance-and-resources/artificial-intelligence-ai-judicial-guidance/>).

¹⁹⁹ La prima parte del documento è dedicata all'"uso responsabile dell'IA (in particolare di quella che interagisce in modo colloquiale) nelle corti e nei tribunali". L'affermazione iniziale dimostra già il requisito che i chatbot costruiti attraverso l'IA dovrebbero essere utilizzati solo da chi è consapevole delle capacità e dei potenziali limiti di queste tecnologie. Sebbene tali limiti siano individuati in tutto il documento, all'inizio ne vengono selezionati alcuni come i "principali": la mancanza di autorità, l'incompleta verificabilità delle informazioni, la scarsa qualità delle risposte, la sovrabbondanza di dati provenienti dal sistema americano.

²⁰⁰ Una sezione specifica è dedicata alla necessità di utilizzare "informazioni accurate e responsabili". Per quanto riguarda il primo aspetto, le informazioni acquisite attraverso uno strumento di IA devono sempre essere verificate prima del loro utilizzo. Gli strumenti di IA possono contenere dati imprecisi, incompleti, fuorvianti o non aggiornati. Anche se affermano di rappresentare una conoscenza giuridica completa, non è necessariamente così. Come si è visto, i LLM possono "allucinare" e inventare casi, citazioni o frasi fittizie, fare riferimento a leggi, articoli o testi giuridici inesistenti, fornire informazioni errate o fuorvianti sulle fonti giuridiche o sulla loro applicazione, o commettere errori di fatto. Uno dei punti più importanti del documento è l'uso corretto delle risposte. Come accennato in precedenza, bisogna essere consapevoli che, data la natura "non supervisionata" di questi strumenti, i dataset di addestramento dell'IA possono riflettere errori e pregiudizi nei dati di addestramento. Questa possibilità deve essere pienamente riconosciuta e ci si deve impegnare per correggere ogni possibile errore.

²⁰¹ Il documento contiene anche una sezione sui fondamenti della "sicurezza informatica". Come è noto, uno dei principali elementi di garanzia contro eventuali abusi informatici è l'utilizzo di apparecchiature certificate e progettate per garantire un elevato livello di sicurezza informatica. Al contrario, l'utilizzo di hardware e software non forniti dal sistema giudiziario può aumentare notevolmente i pericoli o non essere pienamente controllabile dai responsabili della sicurezza informatica nel documento.

²⁰² In un'intervista al Financial Times, il Master of Rolls Geoffrey Vos ha anticipato l'arrivo della Guida sottolineando che le IA non saranno ufficialmente autorizzate ad aiutare il processo decisionale finché la magistratura non sarà assolutamente sicura della bontà di questi strumenti e i cittadini non saranno fiduciosi nel loro utilizzo. A. GRAY, *Judges in England and Wales told to restrict their use of AI in cases*, Financial Times.

²⁰³ Il documento osserva che alcuni tipi di IA sono utilizzati da tempo e senza difficoltà in ambito giudiziario e in molti settori della vita in generale: ad esempio, nei "motori di ricerca per l'autocompilazione delle applicazioni, nei social media per selezionare i contenuti da distribuire, nel riconoscimento delle immagini e nel testo predittivo". Rispetto a questi usi, tutti i professionisti che lavorano in ambito legale hanno una specifica responsabilità per il materiale che producono in tribunale e uno specifico obbligo professionale di garantirne l'accuratezza e l'adeguatezza, come abbiamo visto nel caso di New York e nella decisione del tribunale del Texas. Anche in questi casi, se un'AI viene utilizzata in modo responsabile, si ritiene che non vi sia motivo per un professionista di fare riferimento al suo utilizzo. Tuttavia, si suggerisce che fino a quando la professione legale non avrà acquisito familiarità con queste nuove tecnologie, "gli avvocati potrebbero aver bisogno di ricordare i loro obblighi e di confermare di aver verificato in modo indipendente l'accuratezza di qualsiasi ricerca o citazione di casi generati con l'assistenza di un chatbot AI". La necessità di un uso consapevole dei chatbot creati con gli LLM deve valere anche per coloro che si trovano in tribunale senza una rappresentanza professionale. In questi casi, quando un paziente si rivolge ai motori di ricerca tradizionali per acquisire informazioni mediche, si suggerisce che l'uso di questi strumenti non sia l'unica fonte di consulenza o assistenza per le parti. I giudici, pertanto, devono esercitare un maggiore controllo per verificare l'accuratezza delle informazioni fornite dalle parti.

L'ultima avvertenza riguarda il possibile utilizzo di materiale falso (testi, immagini e video) e l'ulteriore, e più pervasivo, controllo di veridicità a cui devono essere sottoposti i documenti prodotti in tribunale.

La guida redatta dalle supreme magistrature inglesi e gallesi ci aiuta a individuare un presupposto da cui muovere un ragionamento più ampio sull'uso delle GenAI nei processi. Dato l'impatto delle scoperte tecnologiche dirompenti e la loro diffusione, nessuno Stato può far finta che le GenAI non siano un possibile aiuto e, al contempo, un problema per l'esercizio dei propri poteri²⁰⁴. Questa considerazione generale vale tanto per la garanzia dei servizi pubblici quanto per la sicurezza o l'assistenza sanitaria collettiva, e vale anche per l'ambito della giustizia e della garanzia dei diritti fondamentali.

Il documento parte dall'assunto che le GenAI sono diverse dalle IA che abbiamo usato finora, sia per l'uso che per il potere che esprimono nei confronti del soggetto agente, che potrebbe non essere in grado di comprendere le conseguenze che generano. La magistratura inglese è consapevole di questo problema, infatti, nella Guida si chiede implicitamente non come adattare le GenAI ai processi, ma come istruire i giudici a un loro uso consapevole, ora che tali applicazioni sono liberamente disponibili per tutti.

Anche la Corte di giustizia dell'Unione europea ha pubblicato una "Strategia sull'intelligenza artificiale". La strategia indica che queste tecnologie possono essere sfruttate per raggiungere molteplici obiettivi, concentrandosi su tre aree chiave: migliorare l'efficienza e l'efficacia dei procedimenti giudiziari, migliorare la qualità e la coerenza delle decisioni giudiziarie e aumentare l'accesso e la trasparenza per i cittadini dell'UE. La strategia non si ferma a una semplice analisi "performativa", poiché descrive una serie di "principi" sull'IA per la CGUE, sui rischi e sulle modalità di gestione, nonché alcune indicazioni sulla governance di queste tecnologie per il futuro della giustizia nell'UE.

Già dalla sintesi, la strategia sottolinea che l'IA ha un potenziale significativo per la CGUE, sia consentendo l'automazione di compiti semplici in ambito giudiziario e amministrativo, sia offrendo nuove possibilità per la ricerca giuridica, la traduzione e l'interpretazione, migliorando l'accessibilità e fornendo un approccio modernizzato all'accesso alle informazioni.

Queste tecnologie possono essere sfruttate per raggiungere molteplici obiettivi, concentrandosi su tre aree chiave: migliorare l'efficienza e l'efficacia dei procedimenti giudiziari, migliorare la qualità e la coerenza delle decisioni giudiziarie e aumentare l'accesso e la trasparenza per i cittadini dell'UE.

Anche la Commissione per la democrazia attraverso il diritto presso il Consiglio d'Europa, la c.d. Commissione di Venezia, ha affrontato il tema nell'ultimo aggiornamento della propria "Rule of Law Checklist"²⁰⁵.

La Commissione di Venezia sottolinea che l'intelligenza artificiale solleva questioni complesse che investono due piani distinti. Il *primo* riguarda lo sviluppo e la commercializzazione dei sistemi di IA da parte di attori privati, che devono essere assoggettati a obblighi specifici per garantire il rispetto dello Stato di diritto, dei diritti umani e della democrazia lungo l'intero ciclo di vita dei sistemi. Il *secondo* riguarda le condizioni alle quali le autorità pubbliche possono ricorrere all'IA in modo conforme a tali principi. In entrambi i casi è necessario un quadro di governance che preveda valutazioni preventive di rischio e impatto — con riferimento ai diritti umani, allo Stato di diritto e alla

²⁰⁴ La modernizzazione dell'amministrazione e dei servizi non è solo un fattore da considerare per i propri cittadini, ma anche un elemento che posiziona ogni Stato nella geopolitica mondiale ed è una componente fondamentale della sovranità. Le amministrazioni degli Stati dovranno quindi riorganizzarsi, perché chi controlla la produzione e l'uso dell'IA sarà in grado di plasmare il futuro (politico) globale (politico).

²⁰⁵ European Commission for Democracy Through Law (Venice Commission), *The Updated Rule of Law Checklist*, 16 Dec. 2025 (reperibile in: [https://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-AD\(2025\)002-e](https://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-AD(2025)002-e))

democrazia — e un organismo di controllo indipendente, dotato di adeguate risorse investigative e finanziarie, preposto alla supervisione e alla responsabilità.

Nella nuova Checklist, la Commissione di Venezia indica che l'integrazione delle tecnologie digitali e dell'intelligenza artificiale nel sistema giudiziario debba essere preceduta dalla definizione di un quadro normativo rigoroso e bilanciato. Se da un lato tali tecnologie possono migliorare l'efficienza della giustizia e rafforzare i diritti individuali, dall'altro espongono a rischi concreti: la compromissione del diritto a un processo equo, la riproduzione di pregiudizi sistemici, le violazioni della privacy e le disuguaglianze nell'accesso.

La Commissione distingue nettamente tra i sistemi di IA destinati alla gestione amministrativa e quelli impiegati nelle funzioni giudiziarie vere e proprie: i secondi richiedono cautele maggiori, poiché incidono direttamente sull'esercizio della giurisdizione. In ogni caso, le decisioni finali devono restare in capo al giudice, escludendo qualsiasi automatismo decisionale.

I sistemi adottati devono rispettare i diritti fondamentali — con particolare attenzione alla non discriminazione, alla presunzione di innocenza e al diritto alla difesa — e soddisfare requisiti di trasparenza, robustezza e interoperabilità, essere previamente testati in ambienti regolatori sicuri (*sandbox*) e sottoposti a monitoraggio indipendente. È inoltre necessario garantire rimedi efficaci che consentano di contestare sia la scelta di adottare un sistema, sia la sua progettazione, sia il suo utilizzo nel caso concreto.

Una delle istituzioni internazionali più attive nello studio e nella promozione di un uso responsabile dei sistemi di IA da parte degli operatori giudiziari è l'UNESCO.

Dopo aver pubblicato un “Global Toolkit on the AI and Rule of Law for the Judiciary”²⁰⁶ e aver consegnato i risultati di un'indagine sull'alfabetizzazione dei giudici in materia di IA,²⁰⁷ in cui si è riscontrato che quasi il dieci per cento delle organizzazioni giudiziarie ha emanato linee guida adeguate o corsi di formazione in materia di IA, il programma dell'UNESCO “AI and the Rule of Law”, nell'ambito della “Global Judges Initiative”, ha elaborato delle “Linee guida per l'uso dei sistemi di IA nelle corti e nei tribunali”, prima sotto forma di bozza²⁰⁸ e poi in fase finale²⁰⁹. Il documento mira a “offrire una guida completa alle corti e ai tribunali per garantire che l'impiego delle tecnologie di IA sia in linea con i principi fondamentali della giustizia, dei diritti umani e dello Stato di diritto”.

Le Linee guida evidenziano i benefici e i limiti dell'utilizzo dei sistemi di IA nei processi decisionali. Da un lato, l'IA potrebbe migliorare l'efficienza dei procedimenti giudiziari²¹⁰, rendendo più efficace l'accesso alla giustizia e il diritto a un processo equo in tempi ragionevoli. Dall'altro lato, l'IA potrebbe avere un impatto dannoso a causa degli effetti che abbiamo visto sopra riguardo alle allucinazioni e ai possibili effetti discriminatori²¹¹.

²⁰⁶ UNESCO, *Global Toolkit on AI and the Rule of Law for the Judiciary*, CI/DIT/2023/AIRoL/01, 2023 (reperibile in: <https://unesdoc.unesco.org/ark:/48223/pf0000387331>).

²⁰⁷ UNESCO, *Global Judges' Initiative: Survey on the use of AI systems by judicial operators*, cit.. Sulla questione dell'alfabetizzazione dei giudici si veda anche B. CUSTERS, *A fair trial in complex technology cases: Why courts and judges need a basic understanding of complex technologies*, in *Computer Law & Security Review*, 2024, 105935.

²⁰⁸ UNESCO, *Draft UNESCO Guidelines for the Use of AI Systems in Courts and Tribunals*, 2024 (reperibile in: <https://unesdoc.unesco.org/ark:/48223/pf0000389786>) Le linee guida sono state aperte alla consultazione pubblica in inglese fino al 5 settembre 2024.

²⁰⁹ UNESCO, *Guidelines for the Use of AI Systems in Courts and Tribunals*, cit.

²¹⁰ La bozza di linee guida sottolinea che “questi strumenti possono contribuire a migliorare la qualità del lavoro della magistratura facilitando la gestione dei tribunali, la stesura dei documenti, l'esplorazione di argomenti specifici, l'automazione dei compiti e il supporto ai processi decisionali”. Inoltre, una delle caratteristiche principali dei LLM è che facilitano l'interazione tra computer e utenti attraverso il linguaggio naturale”.

²¹¹ La bozza di linee guida sottolinea che “l'adozione di strumenti difettosi e l'uso negligente dei sistemi di IA da parte del sistema giudiziario possono anche compromettere i diritti umani, come il giusto processo e l'equo processo, l'accesso alla giustizia e a un rimedio efficace, la privacy e la protezione dei dati, l'uguaglianza davanti alla legge e la

Sebbene questi strumenti di IA non possano sostituire un ragionamento giuridico qualificato, perché non hanno una comprensione razionale o contestuale di un problema legale, rappresentano una pluralità di tecnologie che possono potenzialmente trasformare e sconvolgere l'intera pratica del diritto e del giudizio giudiziario.²¹² Il loro uso dovrebbe essere gestito con cautela dai non avvocati.

Le Linee guida seguono un quadro tripartito. La prima sezione fornisce tredici principi, principalmente legati alla necessità di proteggere i diritti fondamentali, che le organizzazioni e i singoli membri della magistratura dovrebbero seguire quando adottano e utilizzano qualsiasi sistema di IA.²¹³ La seconda sezione comprende raccomandazioni e indicazioni per le organizzazioni che fanno parte del sistema giudiziario (organi di governo della magistratura, tribunali e corti). Queste indicazioni riguardano: l'adozione di sistemi per la valutazione dell'impatto degli strumenti di IA; le procedure e gli standard interni (gestione del rischio, governance dei dati, cybersecurity); le sessioni di formazione per i giudici e gli elementi specifici per GenAI. La terza parte contiene indicazioni per i singoli membri della magistratura. In questa parte, c'è una sottosezione specifica sull'uso dei sistemi di IA generativa, in particolare "proteggere i dati personali e confidenziali", "principali usi dei LLM", "inaffidabilità come motori di ricerca e per l'analisi giuridica", "consapevolezza dei limiti e dei rischi dei LLM", "verificare i risultati prima di utilizzarli", "uso trasparente", "integrità" e "responsabilità".

6. Profili costituzionali relativi all'utilizzo dei LLM nei processi

L'inarrestabile sviluppo della GenAI ci avverte che è in atto un cambiamento radicale, con effetti di vasta portata, sia nella pratica sia nella teoria del diritto. Dal punto di vista quantitativo, l'IA funziona in modo diverso rispetto alle reazioni semiautomatiche, nelle quali la mente umana ricorre a euristiche e pregiudizi. Tuttavia, da un punto di vista qualitativo, l'IA non ha fatto, e forse non farà mai, il salto dall'elaborazione dei dati all'uso umano, che implica la capacità di ragionare in modo consapevole, riflessivo e responsabile. L'imitazione dell'intelligenza umana affascina coloro che utilizzano i sistemi generativi. Tuttavia, si deve sempre considerare che i risultati di questi tool derivano da semplici associazioni di parole, generate da un algoritmo ignaro del loro significato o della loro rilevanza.

Questo problema tocca direttamente i punti cruciali che dobbiamo affrontare quando utilizziamo l'IA nelle attività giudiziarie, come il processo: Quale concetto di giustizia avevano in mente gli informatici quando hanno programmato l'algoritmo che gira sotto ChatGPT? Chi è responsabile della selezione dei dati da utilizzare per l'addestramento del modello di IA? Possiamo valutare le operazioni di un algoritmo utilizzando uno specifico concetto di giustizia? Possiamo utilizzare questi strumenti di IA in attività giudiziarie e legali senza conoscere la risposta alle domande precedenti?

L'avvento della GenAI ha segnato un salto di qualità nello studio e nell'applicazione dell'IA alle attività legali. L'IA è diventata sempre più capace di riconoscere ed estrarre informazioni da testi, come leggi, sentenze, contratti, saggi giuridici, ecc. È indubbio che la

non discriminazione, nonché i valori giudiziari come l'imparzialità, l'indipendenza e la responsabilità". Inoltre, l'uso improprio dei sistemi di IA può minare la fiducia della società nel sistema giudiziario".

²¹² M. DAHL, e al., *Large legal fictions: Profiling legal hallucinations in large language models*, 2024 (reperibile in: <https://arxiv.org/abs/2401.01301>)

²¹³ Questi principi riguardano principalmente la "correttezza", la "non discriminazione", la "protezione dei dati personali", la "sicurezza", la "sicurezza delle informazioni", la "consapevolezza", la "trasparenza", la "spiegabilità", la "supervisione umana".

velocità di elaborazione dei testi, l'accuratezza dell'estrazione dei dati e la conseguente formazione di risposte coerenti e significative siano travolgenti.

Come hanno sottolineato alcuni studiosi, le informazioni estratte dall'IA si prestano anche a un'ulteriore elaborazione efficiente, poiché spesso si tratta di dati strutturati²¹⁴. Tali dati strutturati sono prodotti come output dell'estrazione, anche nell'elaborazione del linguaggio naturale, progettata per avvicinarsi alla comprensione delle informazioni come fanno gli esseri umani. Sebbene le macchine non abbiano una vera comprensione delle informazioni che elaborano, eccellono nell'estrarre, interpretare e integrare i dati. Le risposte che generano sono spesso indistinguibili da quelle prodotte dall'uomo e, in alcuni casi, possono addirittura superare in qualità i risultati umani²¹⁵. La GenAI ci sta portando, tuttavia, dall'IA ausiliaria a quella autonoma, rendendo quest'ultima una prospettiva più raggiungibile e accessibile per tutti, soprattutto in ambito legale.

Tenendo conto di questi elementi, comprendiamo ora l'impatto sui principi costituzionali in materia di giustizia in una situazione in cui l'IA è ampiamente utilizzata dai professionisti legali, anche al di fuori del processo, come strumento che prende decisioni, in parte o completamente, in modo autonomo.

La graduale, invisibile ma profonda infiltrazione degli algoritmi nella sfera pubblica e privata della vita ha raggiunto un punto critico.²¹⁶ Ora stanno iniziando ad acquisire "legittimità". Le persone smettono di chiedersi se siano giustificate, partendo dal presupposto che l'efficienza significhi una legittimazione sicura. Molti dei documenti emessi dalle istituzioni di tutto il mondo che abbiamo analizzato sopra considerano l'intrusione dell'IA nella giustizia come irrevocabile e irreversibile.

Si percepisce l'IA come uno strumento da utilizzare normalmente, come in passato era per le macchine da scrivere, i computer e internet. Sembra quasi impossibile separare la tecnologia dall'uomo e abbandonare i vantaggi su cui le persone sono abituate a fare affidamento²¹⁷. Nel sistema giudiziario odierno, i principi di efficienza ed efficacia sembrano iniziare a sostituire la "legittimazione democratica". Pertanto, l'aggiunta dell'IA alla giustizia non può essere considerata neutrale e apolitica. L'IA è intrinsecamente incontrollabile in molti aspetti e può operare meglio degli esseri umani nel rispettare i principi costituzionali in materia di giustizia.

Il dibattito sull'IA non è solo teorico, ma anche pratico²¹⁸.

L'emergere della GenAI rappresenta un profondo cambiamento rispetto al concetto di "giustizia digitale"²¹⁹. L'impiego di tali strumenti comporta un duplice rischio. In primo luogo, quando viene condotto attraverso circuiti automatici, il processo manca di qualità umane intrinseche essenziali per il processo decisionale. Ciò solleva preoccupazioni per la perdita di umanità nel giudizio²²⁰. In secondo luogo, i processi giudiziari sono intrinsecamente ancorati a luoghi fisici specifici, come le aule di tribunale, dove la presenza di un giudice conferisce alla procedura una legittimità insostituibile²²¹. L'autorità

²¹⁴ Y. RAZMETAeva, *Artificial intelligence and the end of justice*, cit., 355.

²¹⁵ J. COLLENETTE, e al., *An explainable approach to deducing outcomes in European Court of Human Rights cases using ADFs*, in *Frontiers in Artificial Intelligence and Applications*, 2020, 21–32.

²¹⁶ G. DE GREGORIO, *The Normative Power of Artificial Intelligence*, in *Ind. J. Global Legal Stud.*, 2, 2023, 55–80, 55.

²¹⁷ A. RACHOVITSA, N. JOHANN, *The human rights implications of the use of AI in the digital welfare state: Lessons learned from the Dutch SyRI case*, in *Human Rights Law Review*, 2, 2022, 1–15.

²¹⁸ S.P. DE SOUZA, *The Spread of Legal Tech Solutionism and the Need for Legal Design*, in *European Journal of Risk Regulation*, 3, 2022, 373–388, 373; L.A. DIMATTEO, e al. (a cura di), *The Cambridge Handbook of Lawyering in the Digital Age*, Cambridge University Press, Cambridge, 2021; D. SZOSTEK, M. ZAŁUCKI (a cura di), *Legal Tech*, Nomos, Baden-Baden, 2021.

²¹⁹ E. LONGO, *Giustizia digitale e Costituzione: riflessioni sulla trasformazione tecnica della funzione giurisdizionale*, cit., 45.

²²⁰ L. BREGGIA, *Sprigionare l'umanità del diritto. Una riflessione in tempi disumani*, in *Riv. trim di dir. proc. civ.*, 4, 2025, 939–954

²²¹ A. GARAPON, J. LASSÈGUE, *Justice digitale: Révolution Graphique et Rupture Anthropologique*, cit., 200.

di un'entità artificiale, a prescindere dalla sua presunta intelligenza, non è sufficiente a svolgere questa funzione umana indispensabile.

Occorre una ricontestualizzazione delle norme costituzionali in risposta alla crescente digitalizzazione dei processi giudiziari e le implicazioni per il mantenimento di un processo “equo” in un sistema giudiziario sempre più dipendente da strumenti tecnologici avanzati²²².

Vediamo quattro aspetti del processo da considerare.

In primo luogo, esiste una distinzione fondamentale tra il processo giudiziario e le altre funzioni dello Stato. Il processo giudiziario si concentra sull'applicazione di norme giuridiche a casi particolari, caratterizzati da una rigorosa aderenza alle regole procedurali stabilite. A differenza delle norme procedurali che regolano le funzioni amministrative e parlamentari, il processo giudiziario si occupa principalmente di risolvere le controversie tra parti opposte. Di conseguenza, le attività in questo campo spesso coinvolgono controversie legali radicate in serie di fatti unici.

In secondo luogo, il processo giudiziario si sostiene attraverso il dialogo e le relazioni, fungendo da meccanismo razionale per l'esercizio dell'autorità pubblica. Questo processo si attualizza attraverso una serie di atti formali volti a facilitare la decisione nell'esercizio della funzione giudiziaria.

In terzo luogo, la funzione giurisdizionale richiede una chiara delimitazione delle giurisdizioni, in particolare per quanto riguarda la natura distinta delle situazioni soggettive tutelabili, siano esse diritti soggettivi o interessi legittimi. Inoltre, l'applicazione di un quadro processuale rigido e universale non garantisce la risoluzione rapida di ogni tipi di controversia e può compromettere l'obiettività della decisione giudiziaria. Pertanto, è indispensabile differenziare le procedure di gestione dei casi in base alle caratteristiche specifiche di ciascuno di essi.

In quarto luogo, il processo giudiziario ha un significato sia ideale che materiale. In senso ideale, è fondamentale distinguere quando un diritto riconosciuto è al contempo tutelato. Dal punto di vista materiale, è fondamentale identificare la manifestazione concreta di un diritto oggettivo che viene fatto valere. L'efficacia della tutela giuridica dipende dall'applicazione della legge nel processo giudiziario e dalla realizzazione degli obiettivi “politici” che la sostengono. Dati gli interessi pubblici sostanziali, il processo non può essere considerato privato.

La formula che racchiude l'essenza di queste garanzie, nel più ampio quadro costituzionale, è quella del “giusto processo”²²³. L'evoluzione di questo concetto ha un significato contemporaneo, in particolare nel contesto dell'aumento della digitalizzazione e dell'integrazione di nuove forme di intelligenza artificiale²²⁴.

Poiché il processo giudiziario è il “fattore da cui dipende il grado di realizzazione concreta dei diritti e la misura della loro applicazione pratica”²²⁵, qualsiasi intervento che

²²² A. PAJNO, *L'uso dell'intelligenza artificiale nel processo tra problemi nuovi e questioni antiche*, in *Biolaw Journal*, 1, 2022, 205–223, 205.

²²³ M. CECCHETTI, *Giusto processo (ad vocem)*, in *Enc. dir.*, V. aggiorn., Giuffrè, Milano, 595–627; P. PIRAS, *Il processo amministrativo e l'innovazione tecnologica. Diritto al giusto processo versus intelligenza artificiale?*, in *Dir. proc. amm.*, 4, 2022, 997–1015.

²²⁴ Vedi *infra* multis M. EBERS, *Automating due process—the promise and challenges of AI-based techniques in consumer online dispute resolution*, in *Frontiers in Civil Justice*, a cura di X. KRAMER, et al., Edward Elgar Publishing, Cheltenham, 2022, 142–168, 142; F. PASQUALE, *Inalienable Due Process in an Age of AI: Limiting the Contractual Creep toward Automated Adjudication*, cit., , 42; J. VILLASENOR, V. FOGGO, *Artificial Intelligence, Due Process, and Criminal Sentencing*, in *Michigan State Law Review*, 2, 2020, 295; D.K. CITRON, *Technological due process*, cit., 1249.

²²⁵ A. SIMONCINI, *La dimensione costituzionale della giustizia predittiva. Riflessioni su intelligenza artificiale e processo*, cit., 408.

introduca elementi tecnologici che incidono sulla tutela dei diritti nel processo deve essere valutato in conformità alla clausola del “giusto processo”²²⁶.

Al giusto processo si aggiungono altri principi, tra cui l’indipendenza e l’imparzialità dei giudici, l’esercizio del potere discrezionale, il diritto a un ricorso effettivo e le garanzie del contraddittorio²²⁷. Il processo penale è regolato da ulteriori principi, come la presunzione di innocenza e il diritto all’appello immediato nei confronti dei provvedimenti che violano la libertà personale.

Un principio fondamentale nell’applicazione delle tecnologie digitali alla giustizia è che le garanzie costituzionali riconosciute agli esseri umani devono trovare applicazione anche quando si utilizzano sistemi di GenAI. Ne consegue che in caso di violazione di tali garanzie nell’uso di tali strumenti, le norme di legge che li autorizzano dovrebbero essere ritenute incostituzionali *in parte qua*.

Il fondamento di questa linea di ragionamento risiede nel principio della tutela giurisdizionale, riconosciuto come uno dei principi supremi del nostro ordinamento costituzionale e intrinsecamente legato a quello della democrazia. Esso garantisce a ogni individuo un giudice e un giudizio in qualsiasi controversia. Inoltre, questi diritti sono fortemente protetti e garantiti a livello europeo dall’articolo 47 della Carta dei diritti fondamentali dell’Unione europea (CDFUE)²²⁸, e dall’articolo 6 della Convenzione europea dei diritti dell’uomo (CEDU).

La caratteristica distintiva della giustizia digitale è offrire una nuova sede di ricorso giudiziario in caso di violazione dei diritti o degli interessi. Ciò è particolarmente pertinente quando si utilizzano strumenti digitali, in particolare quelli che adottano un processo decisionale basato su algoritmi. L’efficacia della tutela giudiziaria in questi contesti dipende dall’equità nella progettazione e nell’applicazione di tali tecnologie.

In questo nuovo panorama tecnologico, l’applicazione pratica e contemporanea del principio costituzionale della difesa dei diritti dipende fondamentalmente dalla garanzia di equità nella progettazione e nell’uso di queste tecnologie.

La dottrina ritiene che il concetto di “trasparenza”, inteso in questo contesto come la correttezza degli algoritmi impiegati per chiarire la situazione, debba essere riconosciuto come principio costituzionale a pieno titolo²²⁹. In un quadro costituzionale fondato sui principi di legalità e tutela dei diritti, è imperativo che qualsiasi entità o individuo che utilizzi strumenti decisionali automatizzati dichiari esplicitamente la propria intenzione di farlo²³⁰. Tale trasparenza è essenziale per garantire che tutte le parti potenzialmente interessate siano pienamente consapevoli dello strumento specifico impiegato nel processo

²²⁶ Anche il “giusto processo” della CEDU e della Carta dei diritti fondamentali dell’UE.

²²⁷ K. TERZIDOU, *The use of artificial intelligence in the judiciary and its compliance with the right to a fair trial*, in *Journal of Judicial Administration*, 3, 2022, 154–168, 154.

²²⁸ M. ALMADA, M. DYMITRUK, *Data protection and judicial automation*, in *Research Handbook on EU Data Protection Law*, a cura di E. KOSTA, et al., Edward Elgar Publishing, London, 2022, 270–289, 282.

²²⁹ La citata dottrina utilizza anche l’espressione “lealtà” come vincolo derivante da tale dovere. Si v. A. SIMONCINI, *La dimensione costituzionale della giustizia predittiva. Riflessioni su intelligenza artificiale e processo*, cit., 406. Il concetto di “lealtà dell’IA” è stato incorporato nell’AI Act, in particolare nei casi in cui si dice che i sistemi di IA “interagiscono direttamente” con gli esseri umani. Nello specifico, l’articolo 50 stabilisce che tali sistemi devono essere progettati e sviluppati per garantire che le persone fisiche siano informate quando interagiscono con un sistema di IA. Sebbene questi due principi non siano identici, possono essere intesi come derivanti da una preoccupazione di fondo comune. Poiché un numero sempre maggiore di decisioni viene delegato alle macchine anziché agli esseri umani, è imperativo garantire che questo cambiamento sia percepito come equo.

²³⁰ Questa interpretazione ricorda l’argomentazione di W. HARTZOG, N. RICHARDS, *The Surprising Virtues of Data Loyalty*, in *Emory LJ*, 5, 2021, 985–1034, 985.

decisionale²³¹. Stabilire se una decisione è stata presa tramite processi automatizzati può aiutarci a determinare come interagire con la macchina in modo libero e informale²³².

Con l'avvento dell'IA generativa, la "trasparenza" assume un'importanza ancora maggiore, soprattutto negli scenari in cui tale tecnologia viene "delegata" – anche in modo informale – a prendere decisioni autonome.

Oltre alla lealtà digitale, è necessario garantire ulteriori diritti fondamentali. Tra questi, spicca il diritto di ogni individuo a essere informato delle ragioni alla base di una decisione. Tale principio trova fondamento nell'articolo 24 della Costituzione italiana e nella necessità di assicurare la razionalità interna del processo decisionale mediante un elevato grado di "leggibilità" della funzione giudiziaria. A ciò si aggiunge l'articolo 111 della Costituzione italiana, che stabilisce che "tutti i provvedimenti giudiziari devono essere motivati". La motivazione con cui il giudice accompagna le proprie decisioni è essenziale per legittimare la sua autorità e per garantire la responsabilità nei confronti delle parti, degli avvocati e dei colleghi, nel rispetto della legge.

Tale principio si applica sia in generale a tutti gli atti, pubblici o privati, che esercitano un potere - fungendo da presupposto logico per la loro rivedibilità e, di conseguenza, condizionando l'esercizio del diritto costituzionale alla difesa - sia con particolare forza agli atti giudiziari. In tali atti, la motivazione deve essere sempre accessibile, soprattutto per le decisioni di contenuto decisivo²³³.

L'obbligo di fornire una motivazione deve essere inteso sia come una salvaguardia e un canale per l'autonomia interna del giudice, sia come un elemento essenziale attraverso il quale deve manifestarsi l'esercizio della discrezionalità nell'identificazione, nell'interpretazione e nell'applicazione della legge. Giudicare bene, quindi, significa articolare la motivazione dell'esercizio del potere entro i confini stabiliti dalla legge, indicando chiaramente tutti i fattori che hanno influenzato la decisione finale del giudice.

Coloro che impiegano tecnologie decisionali potenzialmente in grado di influire sui diritti e sulle libertà hanno la responsabilità di farne conoscere l'uso e di garantire che il linguaggio utilizzato sia accessibile agli esseri umani. Questo è un prerequisito per garantire tali diritti e libertà, siano essi giudiziari o di altro tipo.

La convergenza di questi principi porta a una conclusione chiara: solo le tecnologie "spiegabili" o, meglio, comprensibili dovrebbero essere ammesse nelle decisioni che culminano in provvedimenti giudiziari. Di conseguenza, ogni giudice deve comprendere e condividere l'output della macchina²³⁴.

Indubbiamente, una motivazione ben articolata consente di verificare la legittimità di un atto, la sua ragionevolezza, proporzionalità e adeguatezza, nonché di trasmettere una ragione intelligibile per essere considerati atti giuridicamente validi, a prescindere dalla loro potenziale impugnabilità.

Poiché ci affidiamo sempre più ai sistemi di intelligenza artificiale nel processo decisionale, potrebbero verificarsi casi in cui i confini tra il lecito e l'inammissibile

²³¹ Il principio della trasparenza tecnologica come forma di equità algoritmica è intrinsecamente legato al più ampio rispetto della legge. Quando si rivolge alla comunità legale, il giudice deve considerare attentamente le argomentazioni presentate e sforzarsi di garantire che tutti i membri della comunità accettino e rispettino la legge.

²³² L'applicazione, da parte dell'ordinamento giuridico, di principi e regole legali a tali atti è subordinata alla consapevolezza che una decisione è stata presa automaticamente. Ad esempio, l'articolo 22 del Regolamento generale sulla protezione dei dati (GDPR), che riguarda il processo decisionale automatizzato, è stato recentemente interpretato in modo ampio dalla Corte di giustizia dell'Unione europea. A. SIMONCINI, *Art. 22, in Codice della privacy e data protection*, a cura di R. D'ORAZIO, et al., Giuffrè, Milano, 2021, 378–390.

²³³ Inoltre, la motivazione costituisce il meccanismo attraverso cui un giudice stabilisce la legittimità, dimostrando l'aderenza alla legge all'interno del sistema giudiziario più ampio, piuttosto che nell'ambito di un singolo processo. Questa responsabilità è rivolta non solo alle parti e agli avvocati coinvolti, ma anche ai colleghi giudici. N. ZANON, F. BIONDI, *Il sistema costituzionale della magistratura*, cit.

²³⁴ R. CROTOF, *"Cyborg Justice" and the Risk of Technological-Legal Lock-in*, cit., 233.

diventino labili a causa delle caratteristiche strutturali descritte in precedenza. È possibile che si creino zone in cui gli apparati tecnologici, intrinsecamente incomprensibili alla ragione umana, potrebbero sostituire o “catturare” i decisori. Ciò renderebbe estremamente difficile, se non impossibile, verificare l’adeguatezza delle decisioni, la loro legalità e, soprattutto, la loro legittimità.

Il principio di comprensibilità non dovrebbe essere limitato all’esercizio di poteri “pubblici”, dove l’esigenza di trasparenza è particolarmente pressante (oltre all’art. 97 della Cost., è pertinente anche l’articolo 41, comma 2, lettera c, della Carta dei diritti fondamentali dell’UE)²³⁵.

L’interazione tra trasparenza e giustizia dà luogo non solo ai principi di equità e di comprensibilità, ma anche a diverse altre considerazioni significative. In questo contesto, emerge un altro requisito critico, sancito in varie forme nelle costituzioni di molti Stati e nelle carte dei diritti: la necessità che le decisioni siano prese da esseri umani²³⁶.

Nell’interpretazione del nostro sistema giudiziario, questa garanzia è stata articolata in non esclusività, intesa come divieto di affidare le decisioni giudiziarie esclusivamente a sistemi decisionali automatizzati. Una simile pratica limiterebbe indebitamente la libertà interpretativa dei giudici, soggetti “solo” alla legge, una prerogativa costituzionale della funzione giudiziaria. Inoltre, la conoscibilità e l’intelligibilità degli algoritmi dovrebbero costituire garanzie per l’autonomia e l’indipendenza della magistratura. Queste garanzie proteggono la funzione giudiziaria, andando oltre il singolo giudice e includendo la magistratura quale entità collettiva.

Il principio dell’indipendenza funzionale interna esclude che un giudice nel nostro ordinamento sia giuridicamente vincolato da precedenti, con l’unica eccezione dell’obbligo di conformarsi ai principi di diritto stabiliti dalla Corte di Cassazione per i giudici a cui la Corte Suprema rinvia i casi. La Costituzione italiana sottopone i giudici alla legge, obbligandoli ad attenersi a una norma specifica che contempla il caso. Inoltre, i giudici devono valutare le implicazioni della legge in relazione a rapporti, atti e fatti e rispettare le conseguenze legali che la legge stabilisce.

Di conseguenza, le disposizioni costituzionali tutelano non solo gli individui che amministrano la giustizia, ma anche la funzione giudiziaria stessa, comprese le forme procedurali e gli elementi strutturali del processo. I concetti tradizionali di autonomia e indipendenza trovano quindi la loro completa realizzazione nelle caratteristiche che devono definire il processo giudiziario e la sua struttura. Questo quadro è essenziale per sostenere lo Stato di diritto, che richiede, tra l’altro, l’assenza di influenze esterne non derivanti dalla legge, il mantenimento della neutralità, il coinvolgimento di una terza parte imparziale e una struttura procedurale che garantisca che le decisioni siano basate su una valutazione obiettiva dei fatti. Soprattutto, richiede il rispetto dei diritti fondamentali

²³⁵ Questo principio deve invece essere esteso ai soggetti privati, in particolare a quelli che utilizzano questi algoritmi per ottenere un vantaggio significativo o, quantomeno, creare una profonda asimmetria rispetto a singoli soggetti privati o addirittura ad altre persone giuridiche private. Questo principio è riconosciuto dal GDPR (articoli 13 e seguenti e articolo 22) e, più recentemente, dal Digital Services Act (DSA), che impone esplicitamente la trasparenza e la comprensibilità quando si utilizzano algoritmi decisionali, come gli “algoritmi di raccomandazione”, come articolato negli articoli 15, 17 e 27. Una nota importante: è importante distinguere la questione della comprensibilità degli algoritmi decisionali da quella, strettamente correlata e ampiamente dibattuta, della responsabilità civile per i danni causati dagli algoritmi di IA nel diritto privato. La comprensibilità della tecnologia deve essere stabilita come prerequisito prima di determinare la responsabilità, sia in ambito civile che penale.

²³⁶ La *Magna Charta Libertatum* può essere citata tra le prime espressioni di questo principio, ma anche la Costituzione degli Stati Uniti, all’art. III §1 conferisce il potere giudiziario degli Stati Uniti ai tribunali federali e ai loro giudici debitamente nominati. M.R. GROSSMAN, e al., *The GPTJudge: Justice in a Generative AI World*, in *Duke Law & Technology Review*, 1, 2023, 1, 33.

attraverso lo svolgimento di un processo equo, caratterizzato da un contraddittorio tra le parti²³⁷.

Nell'ordinamento giuridico italiano, il concetto di interpretazione giurisprudenziale si è affermato, soprattutto grazie alle decisioni della Corte costituzionale, come una metodologia che trascende la mera applicazione logico-deduttiva dei postulati giuridici. Di conseguenza, l'interpretazione è riconosciuta come un processo intellettuale più complesso che implica decisioni discrezionali e valutazioni autonome da parte dei giudici. Non si tratta semplicemente della "scoperta" del significato di un testo o dell'individuazione dell'approccio ermeneutico più appropriato o di un nuovo e superiore modello di razionalità, ma comporta anche una "ripartizione di poteri". L'interpretazione giudiziaria, quindi, non può essere confinata al solo ambito ermeneutico, ma va compresa nel quadro più ampio del rapporto tra "parole e istituzioni".

Le citate disposizioni costituzionali italiane conferiscono agli organi e ai procedimenti giudiziari caratteristiche essenziali, senza le quali non può esistere né un giudice né l'esercizio della funzione giudiziaria nelle sue varie forme. Tuttavia, i ruoli e i doveri dei giudici sono pienamente articolati nel quadro della legge. Inoltre, è importante riconoscere che l'interpretazione dei testi giuridici è profondamente influenzata dalla struttura della giurisdizione e dalle dinamiche del processo. Allo stesso modo, la decisione di un giudice è influenzata da molti fattori, tra cui considerazioni culturali, politiche, ideologiche e giuridiche, nonché dalla strategia procedurale.

L'implementazione delle tecnologie decisionali porrà probabilmente sfide radicali a questi principi. Come già detto, tali tecnologie generano previsioni future basate su dati esistenti elaborati tramite reti neurali altamente sofisticate. Nel contesto degli strumenti guidati principalmente da algoritmi statistici, il potenziale per un cambiamento delle norme giudiziarie era limitato. Sulla base dei dati utilizzati per addestrare gli algoritmi decisionali, le decisioni prese in passato potevano influenzare in modo significativo gli esiti dei nuovi casi, rendendo difficile cambiare²³⁸.

Tuttavia, in questo scenario, il giudice non sarebbe più libero di essere soggetto solo alla legge, come stabilito dalla Costituzione, ma diventerebbe invece vincolato da una serie di decisioni precedenti, contraddicendo così il principio sancito dal secondo comma dell'articolo 101 della Costituzione italiana. Il precedente stabilito da decisioni giudiziarie precedenti eserciterebbe un'influenza sostanziale e immutabile sulle sentenze future.

Il divieto di qualsiasi forma di predeterminazione delle decisioni giudiziarie, attraverso l'imposizione di vincoli al giudice che vadano oltre l'adeguatezza alle regole stabilite dall'ordinamento giuridico, impone di considerare il potere esercitato da questa tecnologia. L'avvento di nuovi strumenti tecnologici in grado di esercitare un potere autonomo ha portato a profonde trasformazioni nell'amministrazione della giustizia e nell'atto stesso di giudicare. Il processo di digitalizzazione aumenta il potenziale di previsione e di correlazione dei dati, introduce nuove forme di automazione e stabilisce nuove norme tecniche che possono esercitare un'influenza più significativa rispetto alle leggi.

I principi costituzionali sopra delineati precludono inequivocabilmente qualsiasi forma di giustizia "prescrittiva", ossia un sistema che imponga al giudice il contenuto di una decisione²³⁹.

In questo contesto, il parametro costituzionale invocato è quello della ragionevolezza della legge, intesa come assenza di arbitrarietà, come delineata dall'articolo 3 della

²³⁷ J. ULENAERS, *The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?*, in *Asian Journal of Law and Economics*, 2, 2020, 1–38, 1.

²³⁸ Di questo effetto degli algoritmi parla in maniera molto affascinante C. O'NEIL, *Weapons of Math Destruction*, cit., 182 ss.

²³⁹ A.F. DA SILVA, D. MIRANTE, *The Constitutional Limits of Digital Justice*, in *The IT Revolution and its Impact on State, Constitutionalism and Public Law*, a cura di M. BELOV, Bloomsbury Publishing, London, 2021, 199–213, 200.

Costituzione italiana. Questo principio ci permette di affinare ulteriormente l'argomento in esame. Il principio secondo cui il giudice è soggetto solo alla legge presuppone che la legge stessa sia ragionevole, nel senso che non deve limitare arbitrariamente la funzione giudiziaria nella sua applicazione a casi specifici.

L'intelligenza artificiale può supportare il settore della giustizia, non creando un "giudice robot" ma replicando alcune abilità cognitive richieste a giudici e avvocati. Questa capacità può aiutare a mitigare gli errori tipici commessi da questi professionisti.

L'ultimo punto deriva dalle disposizioni collettive degli articoli 3 e 111 della Costituzione italiana, che stabiliscono che "ogni processo si svolge davanti a un giudice terzo e imparziale". L'imparzialità del giudice e la sua posizione di terzo neutrale rispetto alle parti coinvolte sono requisiti fondamentali per un processo equo. Questi requisiti sono essenziali per garantire che il processo sostenga i principi di uguaglianza e di non discriminazione, da un lato, e l'indipendenza della funzione giudiziaria, dall'altro.

È uno sviluppo notevole nel campo della scienza e della tecnologia che gli algoritmi abbiano mostrato tendenze discriminatorie in molti casi, con una frequenza maggiore di quanto si potrebbe aspettarsi²⁴⁰. Questo fenomeno è spesso attribuito a un fattore che nella terminologia tecnica è noto come bias²⁴¹.

Il diritto costituzionale vieta l'uso di algoritmi addestrati su insiemi di dati discriminatori o incompleti, poiché tali algoritmi sono intrinsecamente incapaci di produrre giudizi imparziali.

Indipendentemente dalla traiettoria futura, è imperativo elevare il livello di controllo quando si considera l'integrazione di questi strumenti nella funzione giudiziaria. È fondamentale riconoscere che, in pratica, l'uso diffuso di queste tecnologie nelle professioni legali e quasi legali, in particolare nella redazione di documenti o nella prestazione di consulenze legali, è inevitabile²⁴².

Inoltre, oltre alla necessaria aderenza alla Costituzione italiana e alle leggi, è di fondamentale importanza considerare la coscienza etica e la deontologia professionale di ciascun magistrato. Apponendo la propria firma a una sentenza, il magistrato si assume formalmente e definitivamente la responsabilità dei contenuti in essa espressi, anche se, in realtà, tali contenuti sono stati generati da un sistema di intelligenza artificiale. Tale coscienza etica e deontologia professionale presuppongono un elevato livello di competenza e di formazione nell'uso di tali strumenti.

Purtroppo, il "divario digitale" persiste, soprattutto nel contesto delle disparità culturali ed educative che affliggono il nostro Paese. Dati i potenziali rischi associati a queste tecnologie, tutti gli utenti devono comprenderne a fondo le capacità per evitare danni accidentali o involontari. Questa necessità sottolinea l'urgenza di investimenti più significativi nell'alfabetizzazione digitale, un tema esplicitamente affrontato dall'AI Act con un articolo dedicato (l'articolo 4).

L'integrazione tra uomini e macchine nel determinare i risultati delle azioni pratiche rappresenta un fenomeno molto più ampio e profondo del semplice concetto di giustizia predittiva. Se questa combinazione porterà a un aiuto reciproco, resta incerto. Ciò che è chiaro, tuttavia, è che l'umanità si sta rapidamente avvicinando a un punto critico tra un futuro caratterizzato dalla cooperazione e dalla competizione²⁴³. Pertanto, è essenziale mantenere una prospettiva ampia ed evitare di concentrarsi esclusivamente su aspetti

²⁴⁰ F.J. ZUIDERVEEN BORGESIU, *Strengthening legal protection against discrimination by algorithms and artificial intelligence*, in *The International Journal of Human Rights*, 2020, 1572, 1572.

²⁴¹ M.A. MALEK, *Criminal courts' artificial intelligence: the way it reinforces bias and discrimination*, in *AI and Ethics*, 2022, 233–245, 233.

²⁴² T. SOURDIN, *Robo Justice: Constitutional Issues with Judge AI*, cit., 293.

²⁴³ J. MORISON, T. MCINERNEY, *When should a computer decide? Judicial decision-making in the age of automation, algorithms and generative artificial intelligence*, in *Research handbook on judging and the judiciary*, a cura di S. TURENNE, M. MOUSSA, Edward Elgar Publishing, London, 2024, .

specifici e ristretti, che spesso caratterizzano l'approccio dei giuristi. Questa visione più ampia è necessaria per garantire che le trasformazioni costituzionali in gioco vengano rilevate.

La maggior parte dei principi indicati fa ora parte anche dei documenti sopra esaminati. Questi elementi possono essere considerati una valutazione specifica e iterativa del rischio degli strumenti di IA, prima e durante il loro impiego nelle attività giudiziarie. Naturalmente, questa valutazione non può sostituire la verifica della costituzionalità di tali strumenti.

L'intelligenza artificiale che viene sviluppata in questi settori, così come in altri (penso non solo alla giustizia, ma anche alla salute o al lavoro), ha bisogno di una "costituzione" o, più precisamente, di una legge che definisca le regole per l'uso di un'intelligenza artificiale che agisca esclusivamente per conto dell'umanità e che sia affidabile²⁴⁴.

7. Considerazioni conclusive

In che misura cambierà il ruolo giudiziario in termini di reattività, dato che molti aspetti dell'attività umana, compresi quelli del lavoro degli avvocati e dei giudici, non solo vengono potenziati, ma addirittura sostituiti interamente da tecnologie alternative?

Tali tecnologie stanno già ridefinendo il modo in cui opera la professione legale, con implicazioni per i giudici in virtù delle modalità di preparazione e presentazione dei casi.

Il ruolo del giudice umano, tuttavia, non è semplicemente quello di un elaboratore di dati. Ridurre il ruolo del giudice a tale definizione significherebbe negare non solo l'umanità del giudice, ma anche quella di tutti coloro che si presentano davanti a lui. Una migliore comprensione dell'essenza dell'umanità del giudice contribuirà a garantire che la tecnologia svolga un ruolo appropriato e fondato su principi nel promuovere un sistema giudiziario reattivo. Le intuizioni della psicoanalisi aiuteranno a comprendere questo aspetto e a sviluppare le regole che guideranno le applicazioni future dell'intelligenza artificiale nei processi giudiziari²⁴⁵.

I rapidi progressi dell'IA, esemplificati dagli LLM, segnalano cambiamenti significativi nel diritto e nella professione legale²⁴⁶. Man mano che questi sistemi diventano più sofisticati, è probabile che rimodellino il modo in cui si svolge il lavoro legale e richiedano un approccio ponderato per garantirne l'uso responsabile ed etico²⁴⁷.

Sebbene questi strumenti di IA abbiano acquisito capacità di comprensione, analisi e creazione di bozze di documenti legali senza precedenti rispetto a pochi anni fa, gli operatori del diritto devono utilizzarli con attenzione e responsabilità. Questi sistemi possono talvolta perpetuare inavvertitamente pregiudizi derivanti dai dati di addestramento. Inoltre, il funzionamento interno di questi sistemi è talvolta opaco, rendendo difficile capire perché abbiano prodotto i risultati o il testo che hanno prodotto.

La necessità di maggiore trasparenza e interpretabilità può sollevare preoccupazioni sulla responsabilità e sulla fiducia nel processo decisionale legale.

²⁴⁴ Va ricordato che la storia dell'AI Act è iniziata diversi anni fa con il lavoro del High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, 2019 (reperibile in: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>). Cfr. A. RENDA, *Europe: Toward a Policy Framework for Trustworthy AI*, in *The Oxford Handbook of Ethics of AI*, a cura di M.D. DUBBER, et al., Oxford University Press, Oxford, 2020, 650–666, 6W. HARTZOG, N. RICHARDS, *The Surprising Virtues of Data Loyalty*, cit..

²⁴⁵ T. SOURDIN, R. CORNES, *Do judges need to be human? The implications of technology for responsive judging*, in *The Responsive Judge*, a cura di T. SOURDIN, A. ZARISKI, Springer, Cham, 2018, 87–119.

²⁴⁶ M.R. GROSSMAN, P.W. GRIMM, D. BROWN, M. XU, *The GPTJudge: Justice in a Generative AI World*, cit., 20.

²⁴⁷ J. REGALIA, *From Briefs to Bytes: How Generative AI is Transforming Legal Writing and Practice*, in *Tulsa L. Rev.*, 2-3, 2024, 193–252, 193.

Pertanto, per garantire un uso appropriato delle LLM nel diritto, gli utenti devono sviluppare una comprensione di base del loro funzionamento interno, delle loro capacità e dei loro limiti. Tuttavia, pur tenendo conto di queste preoccupazioni, se utilizzati in modo appropriato, i notevoli progressi dell'IA negli ultimi anni offrono anche la speranza di migliorare la pratica legale e di aumentare l'accesso alla giustizia²⁴⁸.

E allora arriviamo alla domanda cruciale: siamo all'alba di una trasformazione epocale della giustizia?

Se guardiamo alle professioni legali, dobbiamo già riconoscere che la GenAI e la professione legale si integreranno e si svilupperanno, creando forme di carriera complementari all'interno delle professioni legali²⁴⁹; al contempo, nuovi gruppi professionali legali nasceranno da professioni non legali, come le società di "legal tech"²⁵⁰.

Queste due realtà dovrebbero essere regolate dalla legge, dalle regole etiche della scienza e della tecnologia (l'etica dell'intelligenza artificiale) e dalle regole etiche delle professioni legali. Nel giro di pochi anni, saranno necessarie nuove modalità di formulazione delle disposizioni normative e di applicazione delle regole, l'introduzione di nuovi concetti giuridici e la verifica della tenuta di una lunga serie di principi consolidati. Siamo quindi arrivati al punto in cui la connessione fondamentale tra i due campi non può più essere data per scontata, senza sfruttarne appieno ogni aspetto.

La possibilità di utilizzare "macchine" nel diritto ha definitivamente abbandonato la dimensione utopica o distopica della fantascienza per diventare oggetto di dibattito etico, giuridico e politico. Nel prossimo futuro, in seguito alla risoluzione di problemi legati a imprecisioni e pregiudizi, i giudici potrebbero adottare sistemi avanzati di GenAI per semplificare la redazione dei loro pareri.

Al crocevia tra queste riflessioni emerge un altro concetto, quello di "decisore aumentato²⁵¹": una figura giudiziaria che non viene sostituita dalla macchina, né si limita a validarne acriticamente gli output, ma ne integra le capacità analitiche all'interno di un processo deliberativo profondamente umano. Questa prospettiva richiede un ripensamento del ruolo del giudice non come alternativa all'IA, bensì come suo interlocutore critico e responsabile.

Il decisore aumentato si avvale degli strumenti di intelligenza artificiale per accelerare la ricerca giurisprudenziale, analizzare la coerenza degli argomenti, identificare precedenti rilevanti e produrre bozze di motivazione; al contempo, esercita in via esclusiva le funzioni che rimangono irriducibilmente umane: la valutazione del contesto, la ponderazione dei valori in conflitto, la lettura empatica delle parti e — non da ultimo — l'assunzione di responsabilità morale della decisione. Il giudice aumentato non delegherà, ma orchestrerà.

Sul piano epistemologico, questa figura impone una nuova forma di alfabetizzazione giuridico-digitale: il giudice deve saper interrogare criticamente i sistemi di IA, comprenderne i limiti strutturali — inclusi i bias algoritmici e la tendenza all'allucinazione dei modelli generativi — e saper motivare la propria decisione anche laddove l'IA abbia contribuito alla sua formazione. La trasparenza del ragionamento giudiziario non si riduce alla spiegabilità dell'algoritmo: implica la capacità del giudice di rendere conto del percorso decisionale complessivo, incluse le scelte compiute nell'interazione con gli strumenti tecnologici.

²⁴⁸ E. WRIGHT, *Professionals beware: The opportunities and risks of generative ai in legal practice*, in *Intellectual Property Forum: Journal of the Intellectual and Industrial Property Society of Australia and New Zealand*, 2024, 33–41, 33.

²⁴⁹ G. GENTILE, *Human Law, Human Lawyers and the Emerging AI Faith*, 2024, LSE Public Policy Review (reperibile in: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4918632)

²⁵⁰ K. FRAZIER, *The Rise of the Interdisciplinary Lawyer: Defending the Rule of Law in the Age of AI*, in *USFL Rev. F.*, 1, 2023, 19–29, 19.

²⁵¹ A. SANTOSUOSSO, G. SARTOR, *Decidere con l'IA. Intelligenze artificiali e naturali nel diritto*, cit., p. 197 ss.

Dal punto di vista istituzionale, il modello del decisore aumentato pone questioni inedite circa la catena della responsabilità. Se un sistema di GenAI contribuisce alla formazione di un'opinione giudiziaria, chi risponde di un eventuale errore? La risposta non può essere “l'algoritmo”: la responsabilità giudiziaria deve rimanere personale, e ciò esige che il giudice conservi piena signoria sul dispositivo e sulla motivazione. Ne consegue che l'introduzione dell'IA nei processi giudiziari non è un'operazione meramente tecnica, ma richiede una revisione delle garanzie processuali, dei codici deontologici e degli standard formativi della magistratura.

Vi è infine una dimensione antropologica che non può essere trascurata. Il giudizio — nella sua accezione più profonda — è un atto di riconoscimento dell'altro come portatore di diritti, di storie, di dignità. Nessuna macchina, per quanto sofisticata, è in grado di “vedere” la persona che compare davanti al tribunale nel senso pieno che questa visione richiede. Il decisore aumentato non rinuncia a questa responsabilità esistenziale, ma l'esercita con strumenti ancora più potenti, sempre al servizio di una giustizia che rimane, nella sua essenza, umana.

In questo senso, la formula “c'è ancora un giudice a Berlino”, un tempo evocata come metafora della resistenza dell'indipendenza giudiziaria alle pressioni del potere, acquista oggi una valenza ulteriore: c'è ancora un giudice, cioè un essere umano capace di giudizio, anche nell'era dei sistemi intelligenti? E forse è proprio questa persistenza dell'umano, potenziata e non soppressa dalla tecnologia, la sfida più affascinante e urgente che il diritto del ventunesimo secolo è chiamato ad affrontare.