

L'IA nella giustizia: un bilanciamento possibile fra innovazione e democrazia

di *Monica Palmirani*

L'impiego dell'Intelligenza Artificiale (IA) nell'amministrazione della giustizia rappresenta una delle trasformazioni più rilevanti per i sistemi giuridici contemporanei, incidendo sul rapporto tra innovazione tecnologica, autonomia del giudice e garanzie dello Stato di diritto. Il contributo analizza criticamente l'uso dell'IA in ambito giudiziario, evidenziando come l'introduzione di sistemi algoritmici non si limiti a un supporto tecnico, ma influenzi le condizioni stesse dell'esercizio della funzione decisionale. Dopo aver distinto le applicazioni di IA a carattere meramente procedurale da quelle che incidono sull'attività valutativa e interpretativa, l'articolo si concentra sui rischi emergenti legati ai sistemi di IA generativa e agentica, quali opacità, *bias*, variabilità degli *output* e capacità persuasiva. Tali elementi possono compromettere il libero convincimento del giudice, producendo forme di influenza cognitiva difficilmente rilevabili e potenzialmente incompatibili con i principi di indipendenza e responsabilità. Il lavoro approfondisce inoltre il tema del "pensiero artificiale" e delle dinamiche di manipolazione e polarizzazione informativa, mostrando come il problema non consista nella sostituzione dell'operatore umano, bensì nella trasformazione del contesto cognitivo e decisionale in cui il giudice opera. In risposta a tali criticità, viene proposto un approccio di IA ibrida che integri modelli statistici con componenti simboliche e logiche, al fine di rendere il diritto computabile senza irrigidirne la dimensione interpretativa. Infine, l'articolo esamina la catena di responsabilità dell'IA e il ruolo della pubblica amministrazione, sottolineando la necessità di una *governance* strutturata, di requisiti di spiegabilità e di un controllo umano effettivo. La conclusione sostiene che solo un bilanciamento consapevole tra innovazione e prudenza può consentire all'IA di rafforzare, e non indebolire, la legittimità democratica della giustizia.

1. L'Intelligenza Artificiale nella giustizia / 2. Rischi e opportunità verso un bilanciamento / 3. Pensiero artificiale e autonomia del giudice / 4. La legge e la giurisprudenza computabili con l'IA ibrida / 5. La catena di responsabilità dell'IA e il ruolo della pubblica amministrazione / 6. Conclusioni

1. L'Intelligenza Artificiale nella giustizia

L'impiego dell'Intelligenza Artificiale (IA) nel settore della giustizia rappresenta una trasformazione profonda sul piano tecnico, istituzionale e teorico, con

particolare riguardo al rapporto tra i diversi poteri e i cittadini. Non si tratta, infatti, soltanto di introdurre strumenti di automazione, ma di ridefinire il rapporto tra decisione umana, supporto algoritmico e garanzie dello Stato di diritto in uno dei settori più critici per la

tenuta democratica di un paese¹. Le ricerche nello stato dell'arte mostrano che il dibattito ruota intorno a una tensione costante tra efficienza prodotta dall'IA e legittimità, tra innovazione e autonomia del giudice, tra potenzialità tecnologiche e rischi sistemici, tra libero convincimento del giudice e suggerimenti algoritmici.

Si può osservare che ciascun profilo sopra elencato implica un rapporto tra architettura tecnica, garanzie giuridiche e scelte istituzionali. In questa prospettiva, l'analisi non si esaurisce nella descrizione degli strumenti informatici dell'IA, ma coinvolge criteri di affidabilità, *accountability*, auditabilità, spiegabilità e garanzia di controllo umano in tutte le fasi della creazione e messa in operatività degli strumenti di IA.

Per affrontare in modo consapevole questa trasformazione digitale nella giustizia occorre mettere in campo molte competenze e un approccio interdisciplinare, che coinvolge informatica giuridica, teoria del diritto, *governance* delle piattaforme, protezione dei dati, teorie dei sistemi complessi, aspetti tecnologici e organizzazione giudiziaria. Tale intreccio emerge in modo costante già nella prima era di digitalizzazione della giustizia, e dovremmo – specie in Italia che ha visto progetti importanti quali il processo civile telematico e, ora, quello penale – imparare dalle

esperienze pregresse a non ripetere gli stessi errori per mitigare sin da subito le sfide che l'IA ci pone.

Diversi *report* di enti internazionali tratteggiano a livello globale lo stato dell'utilizzo dell'IA nella giustizia (UNESCO², OCSE³) e registrano un aumento dell'uso dell'IA nell'ambito giuridico e della giustizia. Possiamo produrre una tassonomia, rappresentata nella fig. 1.

Il regolamento UE 2024/1689 (AIA) indica che i sistemi di IA utilizzati in ambito giudiziario o nelle attività connesse alla giustizia possono rientrare tra quelli ad alto rischio (vds. AIA, all. III), in particolare quando incidono sui processi decisionali (art. 6, par. 3). Restano invece esclusi da tale qualificazione i sistemi che svolgono esclusivamente funzioni di supporto operativo, amministrativo o procedurale, purché privi di un'influenza sostanziale sulle valutazioni o decisioni degli operatori della giustizia («compito procedurale limitato», art. 6, par. 3, lett. a).

Non tutte le applicazioni nell'ambito della giustizia portano infatti a sbilanciamenti classificati come ad alto rischio. Queste applicazioni includono l'assistenza documentale (correzione, riassunti, traduzioni, confronto tra sentenze, anonimizzazione), compiti amministrativi (assegnazione casi, pianificazione, *trend*, calendarizzazione delle udienze). La scoperta della

ASSISTENTE DOCUMENTALE/INFORMATIVO	SCOPERTA DELLA CONOSCENZA
1. Evitare errori ricorrenti nella redazione delle sentenze (ortografia, grammatica)	10. Riconoscere le connessioni tra i casi
2. Riformulare il testo (tono) o tradurre	11. Correlazioni tra fenomeni (ad esempio, stalking/codice rosso, recidiva/autore del reato, ecc.)
3. Produrre abstract e riassunti della sentenza	
4. Confrontare i testi delle sentenze	
5. Migliorare la ricerca di informazioni legali	SUPPORTO DECISIONALE
6. Anonimizzare/Pseudonimizzare	12. Supporto alla decisione (ODR)
COMPITI AMMINISTRATIVI	13. Esplorare alternative o spiegazioni critiche
7. Pianificare meglio il lavoro di squadra (cooperazione)	14. Verificare l'incoerenza nel testo del giudizio
8. Qualificare i casi per l'assegnazione (ordine del giorno)	15. Classificare i giudizi più pertinenti
9. Fornire tendenze dei casi ricorrenti	

Figura 1 –
Tassonomia
delle funzioni
dell'IA nella giustizia

1. O. Pollicino, *Costituzionalismo digitale. Pensare la democrazia al tempo dell'IA*, Il Mulino, Bologna, 2025.

2. UNESCO: *Guidelines for the Use of Artificial Intelligence Systems in Courts and Tribunals*, Parigi, 2025 (<https://doi.org/10.58338/LIEY8089>); *Draft Guidelines for the Use of AI Systems in Courts and Tribunals*, Parigi, 2025 (<https://unesdoc.unesco.org/ark:/48223/pf0000393682>). In particolare vds., al par. 1, l'elenco degli interessi tutelati:

«Principles: Protection of human rights, Proportionality, Feasibility of benefits, Safety, Information security, Accuracy and reliability, Explainability, Auditability, Transparent and open justice, Awareness and informed use, Responsibility, Accountability and contestability, Human oversight and decision-making, Human-centric and participatory design, Multi-stakeholder governance and collaboration».

3. Nel *report* del 2025 dell'OCSE, *Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions*, su 200 *use case*, 25 sono inerenti alla giustizia, al terzo posto nella scala delle applicazioni. Vds. *Figure 2.1*. «Use cases are most present in public service, civic participation and justice functions».

conoscenza (correlazioni tra fenomeni) e il supporto alla decisione (ODR⁴, alternative decisionali, verifica incoerenze, classificazione precedenti) possono invece nascondere le insidie sopra esposte⁵. Tuttavia, alcune applicazioni sono polifunzionali ed è difficile talvolta dividere questi compiti in modo chirurgico. Sempre più spesso, la parte operativa-amministrativa erode di fatto quella decisionale proprio tramite l'informaticizzazione e l'automazione – quante volte, infatti, il magistrato deve adeguare il proprio lavoro per rispettare le regole di fatto dei sistemi informatici...

La decisione del Tribunale di Torino del 16 settembre 2025⁶ (in cui si afferma che l'utilizzo dell'IA nella redazione degli atti non può prescindere dal controllo critico del professionista così come del giudice, pena la produzione di atti manifestamente infondati) segnala che le questioni sull'IA sono già entrate nel discorso giurisprudenziale. Ciò conferma che il tema non è solo prospettico, ma attuale.

Un elemento di grande rilievo è monitorare gli incidenti⁷ causati dall'IA per comprendere, mediante l'approccio anche empirico, i limiti, i rischi, i tranelli, senza rinunciare al loro utilizzo. Il fenomeno degli incidenti vede una crescita rilevante nel gennaio 2026 con 435 episodi, rispetto al gennaio del 2025 quando erano intorno ai 100⁸. Se è indubbio che anche l'essere umano sbaglia e ha pregiudizi⁹, tuttavia qui ci troviamo davanti al tema di come non rafforzare questi esiti con l'uso dell'IA, ma piuttosto di mitigarli. Si veda, per esempio, il caso di un giudice del Minas Gerais, in Brasile, che ha usato ChatGPT per redigere una sentenza che portava all'assoluzione di un imputato per presunta violenza su una minore¹⁰. Il caso ha sollevato preoccupazioni sull'influenza dell'IA nelle decisioni giudiziarie e sui possibili rischi per i diritti fondamentali¹¹.

La delibera plenaria del Csm dell'8 ottobre 2025 si colloca come cornice istituzionale di riferimento

per evitare questi casi, insistendo su prudenza, consapevolezza, controllo umano per garantire l'indipendenza della magistratura.

2. Rischi e opportunità verso un bilanciamento

L'IA, accanto a innegabili vantaggi che mirano a migliorare l'efficienza e l'efficacia del sistema giustizia, presenta nuovi rischi, che occorre investigare con attenzione per mitigarli. Forse il più insidioso risulta essere la capacità di erodere l'autonomia decisionale¹² del giudice, sovrastandolo di informazioni spesso confermate delle ipotesi iniziali e supportate da argomenti difficilmente verificabili dall'essere umano, a causa di modelli composti da milioni di parametri. Nel contempo, l'uso di una così vasta gamma di informazioni da parte dell'IA generativa fa sì che l'essere umano fatichi a elaborare in modo critico tali dati, se non addirittura ad accedervi per elaborare un pensiero autonomo. Vi sono diverse criticità, specie nelle IA generative, che possono condizionare la capacità di giudizio del giudice, ragion per cui occorre conoscerle per poterle mitigare gli effetti:

a) l'IA generativa produce risultati variabili perché si basa su processi probabilistici e la stessa domanda può avere risposte diverse anche con minime varianti linguistiche, o formulando la stessa domanda in momenti diversi. Questo dipende anche dal tipo di *pre-training* effettuato nei modelli linguistici con grandi parametri. Il 59% dei *dataset* esaminati dal *report* OCSE¹³ sono in inglese, creando così uno sbilanciamento rispetto alle altre lingue;

b) le risposte possono cambiare in base a come viene formulata la richiesta (e.g., *prompt engineering*), influenzando qualità e contenuto dell'*output* a

4. C. Rule, *Online Dispute Resolution and the Future of Justice*, in *Annual Review of Law and Social Science*, vol. 16, 2020, pp. 277-292 (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3711299).

5. M. Palmirani e S. Sapienza (a cura di), *La trasformazione digitale della giustizia nel dialogo tra discipline: Diritto e intelligenza artificiale*, Giuffrè, Milano, 2022.

6. Tribunale di Torino, sez. lav., sent. 16 settembre 2025, n. 2120 (www.giuslavoristi.it/articolo/1913/ia-e-procedimenti-giudiziari-tribunale-di-torino-sentenza-n-2120-del-16-settembre-2025).

7. *Number of reported AI incidents, 2012–25*, in Aa. vv., *AI Index Report 2026*, Stanford University Institute for Human-Centered Artificial Intelligence (HAI), 2026 (https://hai.stanford.edu/assets/files/ai_index_report_2026.pdf).

8. OCSE AIM, 2026 (Chart: 2026 AI Index report).

9. D. Kahneman, *Thinking, Fast and Slow*, Farrar, Straus & Giroux, New York, 2011.

10. <https://nucleo.jor.br/english/2026-03-25-appellate-judge-forgets-ai-prompt/>.

11. <https://oecd.ai/en/incidents/2026-02-24-f1ab>.

12. S. Sapienza, *Decisioni algoritmiche e diritto*, Giuffrè, Milano, 2024.

13. OCSE, *AI in Justice Administration and Access to Justice*, in *Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions*, OCSE, Parigi, 2025, pp. 265 ss.

seconda del prodotto usato (vds. il *report* HAI2026 di Stanford);

c) non è garantita uniformità di qualità, coerenza o affidabilità perché l'*output* può essere impreciso, incompleto o non pertinente (allucinazioni¹⁴, errori nella catena di azione degli *agent AI*¹⁵);

d) i filtri di polarizzazione nel creare *clustering* di sentenze simili possono isolare le sentenze innovative e originali secondo logiche non sempre trasparenti. Si vedano anche le linee guida *Report on Semantic Categorisation of Judicial Decisions in Caselaw Databases with Recommendations* prodotte da TJENI su come è opportuno utilizzare la categorizzazione automatica nelle banche dati giuridiche¹⁶;

e) l'atteggiamento confermativo (*sycophancy*)¹⁷, che viene adottato specie nell'uso dei *chatbot* conversazionali, segue il flusso linguistico dell'utente, tendendo ad assumere un atteggiamento assertivo o in taluni casi anche seduttivo (e.g., Grok) oscurando i punti di debolezza o di fallacie logiche¹⁸ nelle ipotesi formulate dall'utente;

f) l'omissione di informazioni poco frequenti o temporalmente troppo recenti abituano gli operatori a una illusoria completezza dell'informazione, che si traduce nell'arresto ad investigare più approfonditamente;

g) le eccezioni e i casi di ragionamento giuridico complesso sono un punto critico che non tutti gli

strumenti di IA generativa o modelli di IA generale riescono a identificare in modo adeguato, essendo questi strumenti tarati su metodi probabilistici;

h) spesso i sistemi di IA forniscono *output* diversi al variare – anche minimamente – del quesito linguistico dell'utente (*prompt*), rendendo difficile la certezza della risposta;

i) come risulta da recenti studi, i modelli di IA per finalità generali tendono a persuadere gli utenti della correttezza della propria risposta argomentando con dati, fatti e informazioni che l'utente difficilmente riesce a verificare¹⁹;

l) la creatività dei modelli di IA per finalità generali si colloca nella media della creatività umana, rimanendo un ottimo strumento di approssimazione di una realtà media. Questa realtà digitale è ricostruita e riasssemblata dai dati forniti nell'infosfera come somma dell'agire e non agire degli esseri umani e dei dispositivi in dialogo fra loro (sensori e *internet* delle cose). In sintesi, l'IA sembra creativa nella misura in cui colma la nostra non conoscenza, ma non può essere creativa nel significato più autentico²⁰.

Anche il Csm nella sua delibera dell'8 ottobre 2025, approvando le «Raccomandazioni sull'uso dell'intelligenza artificiale», svolge simili considerazioni²¹, ma tratteggia anche importanti raccomandazioni che portano a mitigare tali rischi. Anche la

14. M. Dahl - V. Magesh - M. Suzgun - D.E. Ho, *Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models*, in *Journal of Legal Analysis*, vol. 16, n. 1/2024, pp. 64-93 (<https://doi.org/10.1093/jla/laae003>).

15. A. Rajwade, *Agent AI Infrastructure in Practice: Learn These Key Hurdles to Deploy Production AI Agents Efficiently*, in *Swiss cognitive*, 10 febbraio 2026 (<https://swisscognitive.ch/2026/02/10/agent-ai-infrastructure-in-practice-learn-these-key-hurdles-to-deploy-production-ai-agents-efficiently/>).

16. M. Palmirani, *Semantic Categorisation of Judicial Decisions in the Case law Databases with Recommendations*, Consiglio d'Europa, gennaio 2024 (<https://rm.coe.int/report-on-semantic-categorisation-of-judicial-decisions/1680aeb729>).

17. M. Cheng et al., *Sycophantic AI decreases prosocial intentions and promotes dependence*, in *Science*, vol. 391, n. 6792, 2026 (www.science.org/doi/10.1126/science.aec8352).

18. S. Vagnoni e M. Palmirani, *Fostering Deliberative Democracy in the Digital Age: An Ai-Powered Platform for Enhanced Citizen Engagement in Legislative Processes*, Atti della XI Conferenza internazionale su "eDemocracy & eGovernment" (ICEDEG), Berna, 11-13 giugno 2025, pp. 339-341 (www.computer.org/csdl/proceedings-article/icedeg/2025/11081638/28wXAOqVvXy).

19. A. Kennan - L. Singh - A. García Guevara - M. Ahmed - J. Goodman, *AI-Powered Rules as Code: Experiments with Public Benefits Policy*, Georgetown University et al., 2025 (<https://digitalgovernmenthub.org/publications/ai-powered-rules-as-code-experiments-with-public-benefits-policy-summary/>); F. Salvi - M. Horta Ribeiro - R. Gallotti - R. West, *On the Conversational Persuasiveness of Large Language Models: A Randomized Controlled Trial*, in *arXiv*, 21 marzo 21 2024 (<https://arxiv.org/pdf/2403.14380>).

20. U. Domanti - L. Campidelli - S. Agnoli - A. De Angeli, (2026). *Are Semantic Networks Associated with Idea Originality in Artificial Creativity? A Comparison with Human Agents*, Atti del "CHI 2026" ("Conference on Human Factors in Computing System"), Barcellona, 13-17 aprile 2026, articolo n. 1041, pp. 1-18 (<https://dl.acm.org/doi/proceedings/10.1145/3772318>).

21. Csm, delibera 8 ottobre 2025, «Raccomandazioni sull'uso dell'intelligenza artificiale nell'amministrazione della giustizia» (www.csm.it/portale/web/csm-internet/w/raccomandazioni-intelligenza-artificiale): «Riflessione più articolata deve essere sviluppata con riferimento al tema delle ricerche sulle banche dati giurisprudenziali. L'utilizzo dell'intelligenza artificiale per tale finalità si colloca in un ambito che, sebbene riconducibile a compiti procedurali, può presentare profili di rischio elevati qualora l'*output* generato venga utilizzato come base esclusiva o prevalente nella formazione del convincimento del giudice.

L'IA 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. In questo caso, l'attività si configura come supporto tecnico-organizzativo, riconducibile ai compiti procedurali limitati ai sensi dell'art. 6, par. 3, lett. a) del Regolamento UE 1689/2024. Tuttavia, laddove i sistemi siano progettati per selezionare automaticamente la giurisprudenza "più rilevante", per suggerire orientamenti interpretativi prevalenti o per generare schemi motivazionali basati su *pattern* ricorrenti, si configura un impiego che incide potenzialmente sull'attività valutativa e sull'indirizzo giuridico, e dunque si esce dall'ambito dell'art. 6, par. 3».

Commissione per l'efficienza della giustizia (CEPEJ) del Consiglio d'Europa traccia linee guida per non cadere nel revisionismo e luddismo, ma nel contempo non sottrarsi alla doverosa analisi dei rischi²².

Un bilanciamento, quindi, è possibile mediante appunto integrazione di diverse discipline, una solida *governance*, una cultura della consapevolezza dei rischi che occorre mitigare e monitorare, e non cedendo alle facili soluzioni che spesso il mercato offre come mera trasposizione di altri domini di applicazioni di natura privatistica (industria, logistica, comunicazione).

3. Pensiero artificiale e autonomia del giudice

Il *Global Risks Report 2026* del World Economic Forum²³ ha indicato, fra i primi cinque rischi del prossimo futuro, la manipolazione dell'informazione e la polarizzazione sociale. Nel 2025 la pubblica amministrazione in generale è entrata in una fase completamente nuova, in cui accanto alle applicazioni di IA tradizionale vediamo inserirsi sistemi conversazionali e agentici che hanno la capacità di persuadere gli utenti con metodi personalizzati. Questo processo crea un'illusione di dialogo autentico, che abbassa le difese cognitive e rende il messaggio nella mente dell'essere umano più efficace e duraturo di una lettura di un messaggio statico. L'utente, infatti, tende a percepire le idee condivise e rigenerate dall'IA come proprie e non come imposte dall'esterno e, quindi, ad ancorarle nelle proprie credenze²⁴. Allo stesso tempo, tecniche come il sovraccarico informativo rendono difficile all'essere umano distinguere tra vero e falso, trasformando la persuasione in un processo quasi invisibile²⁵.

Nel contesto della giustizia, se non adeguatamente mitigato, questi fenomeni sono rafforzati dai *chatbot* conversazionali, che hanno la capacità di interagire individualmente con ogni giudice in modo differenziato, creando bolle informative chiuse e polarizzate dalle quali l'utente fatica a uscire e trovare un contraddittorio.

Nonostante i tentativi di regolamentazione, come l'*AI Act* europeo e il DSA²⁶, restano ampie zone grigie difficili da controllare, perché la persuasione può avvenire in modo indiretto e nascosto, sfruttando gli stessi punti di forza e di debolezza degli utenti come leve per il convincimento. In questo modo, la manipolazione diventa impercettibile: la persona è convinta di aver raggiunto autonomamente una conclusione, mentre in realtà è stata accompagnata passo dopo passo lungo un percorso logico già tracciato dall'algoritmo.

Un report pubblicato su *Nature*²⁷ mette in discussione l'idea che le macchine siano intrinsecamente incapaci di convincere. Lo studio, realizzato su larga scala, ha adottato un confronto diretto: da un lato persone reali impegnate a persuadere interlocutori su temi politici sensibili, dall'altro un modello linguistico avanzato coinvolto in conversazioni parallele. I risultati sono netti: l'Intelligenza Artificiale supera gli esseri umani in tutte le principali misure di efficacia persuasiva. La chiave di questo vantaggio è la cd. "personalizzazione dinamica".

Il lavoro pubblicato su *Science*²⁸ introduce un meccanismo ancora più problematico: il cd. "*Gish galloping*" in versione algoritmica. Questa strategia consiste nel travolgere l'interlocutore con una quantità elevatissima di argomentazioni, dati, statistiche e riferimenti, rendendo di fatto impossibile verificarli o contestarli tutti durante la conversazione. Gli studiosi hanno rilevato che i sistemi più efficaci arrivano a utilizzare in media oltre 25 elementi informativi in

22. CEPEJ, *Guidelines on the Use of Generative Artificial Intelligence for Courts*, Strasburgo, dicembre 2025 (<https://rm.coe.int/cepej-2025-18final-en-guidelines-on-the-use-of-generative-ai-for-courts/48802a4ad1>); Id., *Guidelines for the Online Publication of Judicial Decisions and Access to Legal Knowledge*, Strasburgo, novembre 2024 (<https://rm.coe.int/cepej-2024-9-guidelines-on-the-online-publication-of-judicial-decision/1680b2d0de>).

23. World Economic Forum, *Global Risks Report*, Ginevra, 2026 (https://reports.weforum.org/docs/WEF_Global_Risks_Report_2026.pdf).

24. S. Williams-Ceci et al., *Biased AI writing assistants shift users' attitudes on societal issues*, in *Science Advances*, vol. 12, n. 11/2026 (www.science.org/doi/10.1126/sciadv.adw5578).

25. K. Hackenburg et al., *The levers of political persuasion with conversational artificial intelligence*, in *Science*, vol. 390, n. 6777, 2025 (www.science.org/doi/10.1126/science.aea3884).

26. Regolamento (EU) 2022/2065 sui servizi digitali del 19 ottobre 2022, che modifica la direttiva 2000/31/CE (*Digital Services Act*) – <https://eur-lex.europa.eu/eli/reg/2022/2065/oj>.

27. H. Lin - G. Czarnek - B. Lewis - J.P. White - A.J. Berinsky - T. Costello - G. Pennycook, et al., *Persuading Voters Using Human-Artificial Intelligence Dialogues*, in *Nature*, vol. 648, n. 8093, dicembre 2025, pp. 394-401 (https://ideas.repec.org/a/nat/nature/v648y2025i8093d10.1038_s41586-025-09771-9.html).

28. K. Hackenburg et al., *The levers*, op. cit.

appena dieci minuti di dialogo. In un contesto simile non prevale la qualità della verità, ma la quantità di informazioni presentate. La mente umana, sottoposta a questo sovraccarico, tende a cedere per affaticamento cognitivo, finendo per accettare la conclusione proposta dal *chatbot* come plausibile, semplicemente perché sostenuta da un insieme ampio e apparentemente coerente di dati.

I *chatbot* conversazionali non solo cercano di convincere, ma nel fare questo devono prima studiare il proprio interlocutore, definendo il grado di affidabilità dell'essere umano e profilandolo²⁹. Questo significa che non solo i contenuti sono profilati e polarizzati, ma anche i giudici rischiano di essere profilati nell'esercizio delle proprie funzioni, aprendo un tema di grande rilievo per la tenuta democratica.

Se da un lato questi sono strumenti di forte polarizzazione e manipolazione delle opinioni dell'operatore, allo stesso tempo esiste ancora uno spazio di resistenza fondato sulla consapevolezza e sul pensiero critico, speculativo e controfattuale, che possono portare a una positiva collaborazione fra giudici e IA. In questi contesti si possono innescare processi di autoapprendimento per imparare a dialogare con *chatbot* senza farsi manipolare, sviluppare un pensiero speculativo usando il *prompt engineering*, integrando le piattaforme con i cd. agenti "avvocato del diavolo", che cercano di confutare le ipotesi iniziali evidenziando lacune, debolezze logiche, o dati errati.

Il futuro dell'IA nella giustizia dipenderà sempre più dalla capacità collettiva di difendere l'autonomia del pensiero dei suoi operatori di fronte a strumenti sempre più sofisticati di influenza invisibile e, quindi, non di vietare, ma di governare anche attraverso una formazione mirata. Inoltre, strumenti come *Agentic AI* dedicati a questo scopo possono intercettare fallacie logiche negli argomenti degli esseri umani, ma anche nelle conversazioni prodotte dall'IA genera-

tiva. Un agente detto "giudice" potrà fare sintesi fra molti punti di vista e fungere anche da mediatore in una comparazione con diverse ipotesi giuridiche. In questo contesto, le forme di *Adversarial Agentic AI*, *Critic Agent*, *Socratic Agent*, *LLM-as-Judge* possono mitigare la creazione di un pensiero unico.

4. La legge e la giurisprudenza computabili con l'IA ibrida

Oggi si confrontano tre approcci alla normazione computabile:

- *Rules as Code*³⁰: prima si modella la norma in forma tecnica e poi si traduce in testo;
- *Law as Code*: si parte dal testo giuridico e lo si rende successivamente computabile;
- *Digital-Ready Policy*³¹: la normativa è progettata sulla base degli obiettivi di *policy* e dei suoi effetti.

Questi modelli sono complementari e da usare caso per caso, impiegando una metodologia chiamata "IA ibrida"³², facendo attenzione a non cristallizzare in un codice difficilmente modificabile principi giuridici che sono elastici e permeabili a seconda del contesto. Inoltre, anche i sostenitori dell'approccio "code-first" riconoscono che non esistono ancora condizioni tecnologiche tali da renderlo dominante³³.

Nel *report* HAI2026 di Stanford vi sono *benchmarking* di modelli LLM effettuati su giurisprudenza USA e canadese³⁴: Grok4.3 e GPT5.1 sfiorano rispettivamente il 79.31% e il 73.42% di *accuracy*. Non esiste un analogo studio per l'italiano e il sistema giuridico domestico, ma molto più occorrerebbe valutare anche i *task* specifici (traduzione, sommari, ricerche pertinenti, classificazione, *clustering*, regressione, *prediction*, etc.). In particolare, questi *report* non entrano nel dettaglio dei seguenti parametri:

- a) gli LLM tendono a trascurare la semantica e il contesto normativo;

29. V. Lerman e Y. Dover, *A closer look at how large language models 'trust' humans: patterns and biases*, in *Proc. A*, vol. 482, n. 2335, aprile 2026 (<https://doi.org/10.1098>).

30. R. Kennedy, *Rules as Code and the Rule of Law: Ensuring Effective Judicial Review of Administration by Software*, in *Law, Innovation and Technology*, vol. 16, n. 1/2024, pp. 170–193 (<https://doi.org/10.1080/17579961.2024.2313801>). J. Mohun e A. Roberts, *Cracking the Code: Rulemaking for Humans and Machines*, OCSE Working Papers on Public Governance, n. 42, Parigi, 2020 (<https://doi.org/10.1787/3afe6ba5-en>).

31. <https://interoperable-europe.ec.europa.eu/collection/digital-ready-policy-making/digital-ready-policies>.

32. M. Palmirani - S. Sapienza - K. Ashley, *A Hybrid Artificial Intelligence Methodology for Legal Analysis*, in *BioLaw Journal*, n. 3/2024, pp. 389–409 (<https://doi.org/10.15168/2284-4503-3206>); M. Palmirani - M. Corazza - G. Longo - S. Sapienza, *Hybrid AI to Enhance Legal Drafting with LEOS*, Commissione europea, Lussemburgo, dicembre 2025 (<https://doi.org/10.2799/1298773>).

33. O.R. Goodenough e P.J. Carlson, *Words or Code First? Is the Legacy Document or a Code Statement the Better Starting Point for Complexity-Reducing Legal Automation?*, in *Philosophical Transactions A*, The Royal Society, vol. 382, n. 2270, 2024 (<https://royalsocietypublishing.org/rsta/article/382/2270/20230160/112545/Words-or-code-first-is-the-legacy-document-or-a>).

34. www.vals.ai/benchmarks/case_law_v2. https://www.vals.ai/benchmarks/legal_bench.

b) non gestiscono adeguatamente le citazioni e le relazioni tra norme;

c) confondono le dimensioni temporali rilevanti (come abrogazioni o vigenza);

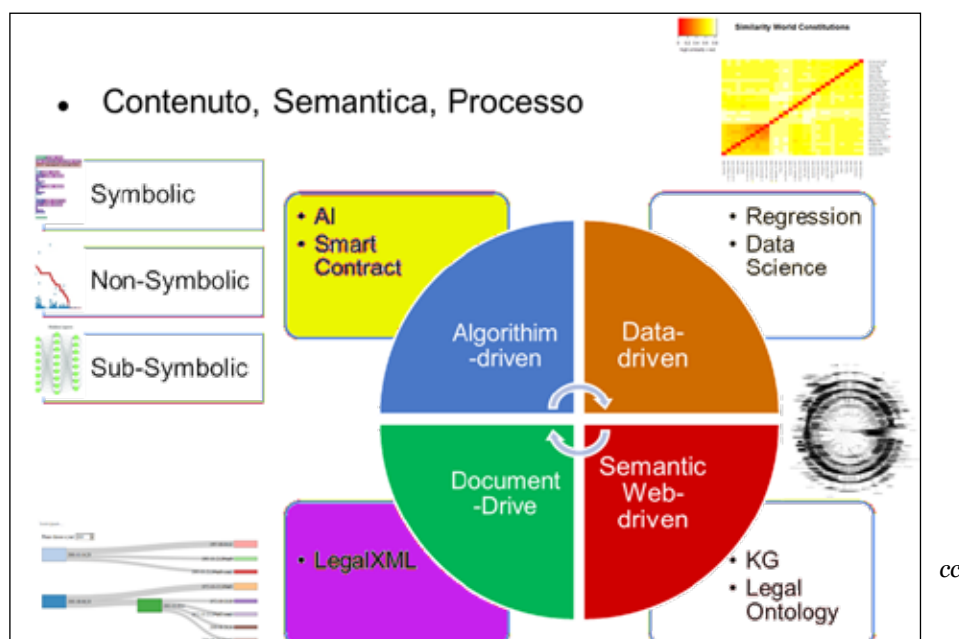
d) faticano con la logica e il linguaggio tecnico-giuridico e non incorporano regole implicite dell'ordinamento (e.g., gerarchie delle fonti, relazione con la Corte costituzionale).

Ne deriva una capacità limitata di interpretare correttamente testi legislativi con molte modifiche nel tempo o casi giuridici complessi.

Per superare questi limiti è preferibile adottare un approccio ibrido, che integri modelli statistici con IA simbolica, *web* semantico e logica giuridica, anche attraverso *standard* come *Akoma Ntoso*³⁵. In questo contesto, gli *Agentic AI* offrono ulteriori possibilità³⁶, permettendo di suddividere compiti complessi tra agenti specializzati e di gestire processi di ragionamento grazie a meccanismi di memoria integrata da regole deterministiche.

Large Language Models (LLM) offrono capacità avanzate, ma pongono problemi di opacità sul percorso adottato per giungere alla risposta a fronte di un quesito. Il tema della *explicability*, richiamato in molte linee guida (UNESCO, OCSE, CEPEJ, HAI),

assume quindi rilievo come requisito di comprensibilità, verificabilità e controllo della correttezza dell'esito, nonché condizione essenziale per assicurare l'autonomia del giudice. In ambito giuridico, la spiegabilità è connessa alla motivazione, alla contestabilità e alla responsabilità, al libero convincimento. Spesso si parla di diritto alla spiegazione (vds. anche GDPR, art. 22); tuttavia, la spiegazione è un oggetto complesso, che implica diversi componenti: comunicante, ricevente, comunicato. A questi elementi si devono aggiungere altri fattori, quali le capacità cognitive e le competenze del ricevente, la forma della comunicazione (testuale, visuale, supportata da dati), le motivazioni del comunicante, lo scopo della comunicazione, la capacità ricettiva del ricevente, le condizioni (temporali, ambientali, spaziali). Per questo motivo è meglio utilizzare il termine "spiegabilità"³⁷ come capacità di produrre una spiegazione, semmai con una interazione persona-macchina tale da inquadrare correttamente i parametri della spiegazione attesa. I sistemi di IA applicati alle attività della giustizia dovrebbero essere integrati da *Agentic AI* dedicati a verificare il rispetto delle regole di *safeguards*, moduli dedicati alla spiegabilità e a verifiche controfattuali³⁸, al fine di evidenziare



35. *LegalDocML (Akoma Ntoso) Version 1.0*, OASIS (Organization for the Advancement of Structured Information Standards) Burlington, Massachusetts, 2018 (<https://docs.oasis-open.org/legaldocml/akn-core/v1.0/akn-core-v1.0.html>).

36. L. Floridi, *La differenza fondamentale. Artificial Agency: una nuova filosofia dell'intelligenza artificiale*, Mondadori, Milano, 2025.

37. S. Sapienza e M. Palmirani, *Algorithmic Knowability: A Unified Approach to Explanations in the AI Act*, in R. Guidotti - U. Schmid - L. Longo (a cura di), *Explainable Artificial Intelligence*, atti della Conferenza xAI, Istanbul, 9-11 luglio 2025, in *Communications in Computer and Information Science*, vol. 2576, Springer, 2026 (https://doi.org/10.1007/978-3-032-08317-3_9); *Ibid.*, *Big Data, Explanations and Knowability*, in *Ragion Pratica*, n. 2/2021, pp. 349-364.

38. Il pensiero controfattuale è stato evocato dal Senato della Repubblica come strumento per esercitare il pensiero critico nei confronti del dialogo con l'IA (www.senato.it/application/xmanager/projects/leg19/file/repository/UVI/strumenti/Strumenti_1.pdf).

lacune nella robustezza del ragionamento giuridico esito dell'elaborazione. Le categorie del ragionamento dell'essere umano seguono processi diversi³⁹ e non si possono paragonare, ma si possono integrare se l'utente finale riesce a ottenere informazioni sufficienti a crearsi un'opinione autonoma.

5. La catena di responsabilità dell'IA e il ruolo della pubblica amministrazione

Non tutto, però, è delegato alla tecnica o alla sorveglianza umana dell'utente finale. Occorre una *policy* di sorveglianza della catena di responsabilità, oltre alla già citata *governance* dell'IA.

Se nel privato emerge l'urgenza di definire la catena delle responsabilità fra fornitore e utente, nell'ambito della giustizia, sia come istituzione pubblica

dell'ordinamento giudiziario (giudici) sia come ente della pubblica amministrazione (Ministero della giustizia), occorre identificare il confine dell'azione che l'organismo attua per raffinare strumenti messi in opera dai fornitori (modello di IA per finalità generali o di rischio sistemico)⁴⁰, che utilizzano per esempio strumenti di intelligenza artificiale generativa generalistici. La regolazione europea e le linee guida collegate⁴¹ in tema di IA nell'art. 25⁴² spostano, infatti, il ruolo di semplice utilizzatore a fornitore di nuove soluzioni se l'intervento di personalizzazione apporta «un cambiamento significativo nel modello», ovvero se supera $1/3$ di 10^{23} *Flops* di potenza di calcolo (compute), oppure se viene cambiato lo scopo passando a uno scenario ad elevato rischio, oppure se viene apposto il proprio marchio indipendentemente dagli altri parametri elencati su una soluzione di modello di IA per finalità generali.

Tipo di modifica	Potenza di computazione utilizzata	Esito
<i>Prompt tuning</i> , RAG, integrazione	Minimo	Rimane <i>deployer</i>
<i>Fine-tuning</i> leggero (LoRA, <i>adapter</i>)	Solitamente $< 1/3$ di 10^{23} <i>Flops</i>	Rimane <i>deployer</i>
<i>Fine-tuning</i> massiccio o <i>re-training</i> parziale	$> 1/3$ di 10^{23} <i>Flops</i>	Diventa <i>provider</i> GPAI
<i>Fine-tuning</i> massiccio o <i>re-training</i> parziale	$> 10^{25}$ <i>Flops</i>	Diventa <i>provider</i> GPAI a rischio sistemico
Cambio scopo → <i>high-risk</i>	Indipendente dalla capacità di calcolo	Diventa <i>provider high risk</i>
<i>Rebranding</i> + immissione sul mercato	In ogni caso	Il <i>deployer</i> diventa <i>provider</i>

Tabella 1 – Tabella di confronto rispetto al tipo di agente, tipo di applicazione, potenza di calcolo

Tale distinzione risulta determinante ai fini della corretta individuazione degli obblighi che il Ministero della giustizia è tenuto ad assolvere in qualità di utilizzatore di sistemi di IA ad alto rischio, nonché per definire le specifiche esigenze di conformità da richiedere all'intera catena di fornitura coinvolta.

Quest'ultimpuò articolarsi in una pluralità di soggetti, frequentemente qualificati, spesso in modo non esplicito, come “fornitori a valle”⁴³ o, con terminologia di settore, come integratori di sistemi di intelligenza artificiale. Ricordiamo che l'*AI Act* elenca fra le applicazioni ad alto rischio anche le ricerche

39. L.F. Barrett e E.K. Miller, *Categorization is 'baked' into the brain*, in *Nature Reviews Neuroscience*, 13 aprile 2026 (<https://doi.org/10.1038/s41583-026-01036-2>).

40. Vds. art. 3 (definizioni dell'*AI Act*), spec. pt. 63 ss.

41. Commissione europea, *Commission Guidelines on the Scope of the Obligations for Providers of General-Purpose AI Models Established by Regulation (EU) 2024/1689 (AI Act)*, C(2025) 7719 final, Bruxelles, 19 novembre 2025.

42. L'art. 25 dell'*EU AI Act* (regolamento UE 2024/1689) disciplina le «Responsabilità lungo la catena del valore dell'IA».

43. Vds. art. 3, par. 1, n. 65, reg. (UE) 2024/1689 (*AI Act*). Nella versione in lingua inglese il medesimo attore è definito «*downstream provider*».

documentali delle banche dati giuridiche (legislative e giurisprudenziali)⁴⁴ che possono in qualche misura influenzare il libero convincimento del giudice, escludendo quelle applicazioni che siano meramente funzionali ai compiti procedurali (e.g., calendari, calcoli tariffe). Tuttavia, questa materia potrebbe rientrare nel supporto tecnico-organizzativo *ex art.* 6, par. 3, lett. *a* dell'IA, ma ne esce quando i sistemi influenzano la selezione della giurisprudenza o l'interpretazione giuridica della legge (e.g., selezione della rilevanza delle leggi pertinenti a un caso), incidendo sulla funzione valutativa dei magistrati.

Ad ogni buon conto, emerge un ruolo che il Ministero della giustizia deve svolgere, ossia proteggere l'istituzione democratica da un'intrusione troppo aggressiva di sistemi eterogeni oscurativi dell'informazione che può servire ai suoi operatori di giustizia. Esso deve, altresì, definire dei *limiti* alle applicazioni senza una adeguata personalizzazione rispetto al nostro sistema giustizia e normativo italiano. Spesso, i *provider leader* nell'IA possono introdurre, anche involontariamente, un'impostazione metodologica e di contenuti provenienti da logiche di *pre-training* che non appartengono alla nostra cultura giuridica (e.g., *common law*, *predictive* basato sul precedente vincolante, concetti giuridici linguisticamente distanti dalla nostra tradizione giuridica), inquinando così la neutralità dello strumento sin dall'origine. Questo può riguardare anche l'introduzione di *bias*⁴⁵ estranei alla nostra cultura, che possono tuttavia trattare in modo non bilanciato e ugualitario gruppi di cittadini rispetto a genere, etnia, lingua, condizione economica.

La letteratura mostra che, per ridurre i *bias* discriminatori dovuti a dati etero-introdotti e non ben partizionati, non sono sufficienti strumenti tecnologici, ma è fondamentale affiancare l'esperienza umana, la sorveglianza di esperti di dominio giuridico, regole

di *governance* costante⁴⁶ e una *policy* chiara di *design* dei sistemi di IA che tenga conto del contesto e della tradizione giuridica.

6. Conclusioni

L'intelligenza artificiale può contribuire a rendere i sistemi giudiziari più efficienti, accelerando l'analisi dei documenti e riducendo i tempi di gestione dei procedimenti. Può, inoltre, migliorare l'accesso alla giustizia, facilitando l'orientamento dei cittadini e la comprensione delle procedure attraverso strumenti di supporto automatizzato. Infine, se adeguatamente governata, può rafforzare la coerenza delle decisioni e il supporto alle attività interpretative, riducendo carichi ripetitivi e valorizzando il ruolo del giudice nelle valutazioni più complesse.

Tuttavia, è indubbio che strumenti di IA, specie quelli GPAI o a rischio sistemico, applicati al mondo della giustizia necessitano di un supplemento di *governance*, di una *policy* strutturata, di processi di monitoraggio e sorveglianza, di *design* specifico personalizzato, adattando strumenti di mercato al linguaggio giuridico, alla cultura di dominio, alle esigenze dei magistrati e della pubblica amministrazione.

Questo presidio non deve cedere sotto le spinte efficientiste, sotto la pressione dei *vendor* e nell'illusione di facili risultati che promettono di saltare passaggi essenziali quali formazione, sperimentazione, valutazione dei risultati, mitigazione dei rischi e dei *bias*. Solo competenze interdisciplinari che abbraccino diverse competenze possono raggiungere risultati duraturi, che non risentano delle mode del momento. L'entusiasmo deve essere compensato da un bilanciamento fra innovazione e prudenza, nel rispetto dei diritti fondamentali, dei principi costituzionali e delle istituzioni democratiche.

44. L'allegato III include, nel pt 5: «Accesso a servizi privati essenziali e a prestazioni e servizi pubblici essenziali e fruizione degli stessi» e, nel pt. 8, lett. *a*: «Amministrazione della giustizia e processi democratici:

a) i sistemi di IA destinati a essere usati da un'autorità giudiziaria o per suo conto per assistere un'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti, o a essere utilizzati in modo analogo nella risoluzione alternativa delle controversie».

45. A.S. Almasoud e J.A. Idowu, *Algorithmic fairness in predictive policing*, in *AI Ethics*, n. 5/2025, pp. 2323-2337 (<https://doi.org/10.1007/s43681-024-00541-3>).

46. F. Lagioia - R. Rovatti - G. Sartor, *Algorithmic Fairness through Group Parities? The Case of COMPAS-SAPMOC*, in *AI & Society*, vol. 38, n. 2/2023, pp. 459-478; I. Cofone e W. Khern-am-nuai, *The Overstated Cost of AI Fairness in Criminal Justice*, in *Indiana Law Journal*, vol. 100, n. 4/2025, pp. 1431-1478 (www.repository.law.indiana.edu/cgi/viewcontent.cgi?article=11578&context=ilj).