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

di Monica Palmirani

professoressa ordinaria di Informatica Giuridica Università di Bologna, Capo dell'Unità di ricerca "AI for Law and Governance", Presidente dell'International Association for Artificial Intelligence and Law

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.

Intelligenza artificiale; amministrazione della giustizia; autonomia del giudice; AI Act; governance; spiegabilità; IA ibrida.

The use of artificial intelligence (AI) in the administration of justice represents one of the most significant transformations affecting contemporary legal systems, reshaping the relationship between technological innovation, judicial autonomy, and the guarantees of the rule of law. This article offers a critical analysis of AI adoption in the judicial domain, arguing that algorithmic tools do not merely provide technical support but actively influence the conditions under which judicial decision-making is exercised. After distinguishing between procedural AI applications and those that directly affect evaluative and interpretative activities, the paper focuses on the risks associated with generative and agentic AI systems. Particular attention is paid to issues of opacity, bias, output variability, and persuasive capacity, which may undermine judges' independent reasoning and introduce subtle forms of cognitive influence incompatible with judicial independence and accountability. The contribution further explores the notion of "artificial thinking" and the mechanisms of information manipulation and polarization enabled by conversational systems. In this perspective, the central challenge is not the replacement of human decision-makers, but the transformation of the cognitive and deliberative environment in which judges operate. To address these challenges, the article advocates a hybrid AI approach that combines statistical models with symbolic and logic-based components, enabling computable law while preserving interpretative flexibility. Explainability and verifiability are identified as essential requirements to maintain meaningful human oversight. Finally, the paper examines the AI responsibility chain and the role of public administration, emphasizing the need for structured

governance frameworks, transparency obligations, and context-sensitive design choices. It concludes that only a careful balance between innovation and prudence can ensure that AI strengthens, rather than weakens, democratic legitimacy in judicial systems.

Artificial intelligence; justice administration; judicial autonomy; AI Act; governance; explainability; hybrid AI.

Sommario: 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 paese[1]. 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 (UNESCO[2], OECD[3]) e registrano un aumento dell'uso dell'IA nell'ambito giuridico e della giustizia. Possiamo produrre una tassonomia, rappresentata nella fig. 1.

---

[1] Oreste Pollicino, Costituzionalismo digitale. Pensare la democrazia al tempo dell'IA, Bologna, Il Mulino, 2025.

[2] UNESCO. Guidelines for the Use of Artificial Intelligence Systems in Courts and Tribunals. Paris: UNESCO, 2025. https://doi.org/10.58338/LIEY8089; UNESCO. Draft Guidelines for the Use of AI Systems in Courts and Tribunals. Paris: UNESCO, May 2025, https://unesdoc.unesco.org/ark:/48223/pf0000393682. In particolare si veda il paragrafo 1 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'OECD. *Governing with Artificial Intelligence: AI in Justice Administration and Access to Justice* su 200 use case 25 sono inerenti la giustizia, al terzo posto nella scala delle applicazioni. Vedi Figure 2.1. Use cases are most present in public service, civic participation and justice functions.

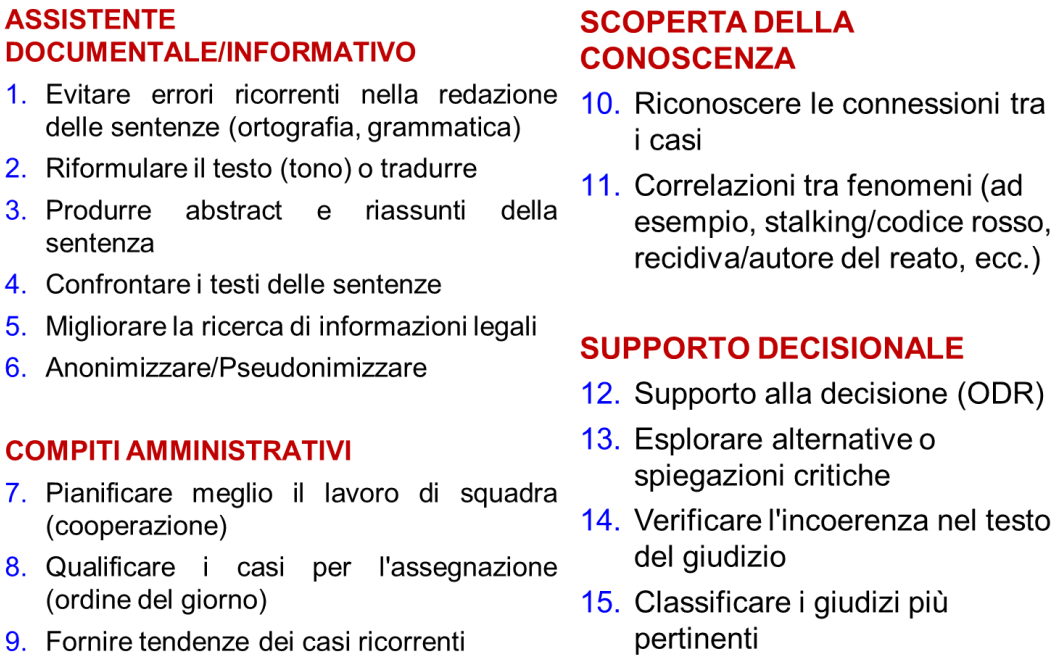

Figura 1 – Tassonomia delle funzioni dell'IA nella giustizia.

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 (si veda allegato III dell'AIA), 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 altro 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 conoscenza (correlazioni tra fenomeni) e il supporto alla decisione (ODR[4], alternative decisionali, verifica incoerenze, classificazione precedenti) possono invece nascondere le insidie sopra esposte[5]. Tuttavia, alcune applicazioni sono polifunzionali ed è difficile talvolta dividere questi compiti in modo chirurgico e sempre più spesso la parte operazionale-amministrativa erode di fatto quella decisionale proprio tramite l'informatizzazione 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 2025 [6] - in cui si afferma che l'utilizzo dell'intelligenza artificiale 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

[4] Online Dispute Resolution. Rule, Colin. "Generative AI and Online Dispute Resolution: Opportunities and Risks." *International Journal of Online Dispute Resolution* 12, no. 1 (2025): 1–28.
[5] Palmirani, Monica, Salvatore Sapienza (eds.). *La trasformazione digitale della giustizia nel dialogo tra discipline: Diritto e intelligenza artificiale*. Milano: Giuffrè Francis Lefebvre, 2023.
[6] Tribunale di Torino, sezione Lavoro, sentenza n. 2120 del 16 settembre 2025, in https://www.giuslavoristi.it/articolo/1913/ia-e-procedimenti-giudiziari-tribunale-di-torino-sentenza-n-2120-del-16-settembre-2025

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 incidenti[7] 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 100[8]. Se è indubbio che anche l'essere umano sbaglia e ha pregiudizi[9], 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 minore[10]. Il caso ha sollevato preoccupazioni sull'influenza dell'IA nelle decisioni giudiziarie e sui possibili rischi per i diritti fondamentali[11].

La Delibera plenaria dell'8 ottobre 2025 del Consiglio Superiore della Magistratura - CSM 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 ad 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 decisionale[12] del giudice sovrastandolo di informazioni spesso confermative 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 fatica 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 e per questo occorre conoscerle per poterne 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 OECD[13] sono in inglese creando 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 seconda del prodotto usato (si veda il report HAI2026 di Stanford);

---

[7] Number of reported AI incidents, 2012–25, in the Stanford Institute for Human-Centered Artificial Intelligence (Stanford HAI), *AI Index Report 2026* (Stanford University, 2026), https://hai.stanford.edu/assets/files/ai_index_report_2026.pdf.
[8] OECD AIM, 2026 | Chart: 2026 AI Index report.
[9] Kahneman, Daniel. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011.
[10] Vedi https://nucleo.jor.br/english/2026-03-25-appelate-judge-forgets-ai-prompt/
[11] https://oecd.ai/en/incidents/2026-02-24-f1ab.
[12] Sapienza, Salvatore. *Decisioni algoritmiche e diritto*. Milano: Giuffrè, 2024.
[13] OECD. Governing with Artificial Intelligence: AI in Justice Administration and Access to Justice. In Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions. Paris: OECD Publishing, 2025. Figure 1.4. More than half (59%) of open-source AI training datasets are in English in https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/06/governing-with-artificial-intelligence_398fa287/795de142-en.pdf

c) non è garantita uniformità di qualità, coerenza o affidabilità perché l'output può essere impreciso, incompleto o non pertinente (allucinazioni[14], errori nella catena di azione degli agentic 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 giuridiche[16];

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 logiche[18] nelle ipotesi formulate dall'utente;

f) l'omissione di informazioni poco frequenti o temporalmente troppo recenti abituano gli operatori ad una illusoria completezza dell'informazione che si traduce nell'arresto a 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) 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 verificare[19];

l) la creatività dei modelli di IA per finalità generali si collocano nella media della creatività umana rimanendo un ottimo strumento di approssimazione di una realtà media. Questa realtà digitale è ricostruita e riassemblata dai dati forniti nell'infosfera come somma dell'agire e il 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 autentico[20].

---

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

[15] Rajwade, A. Agentic AI Infrastructure in Practice: Learn These Key Hurdles to Deploy Production AI Agents Efficiently. Google Research, 2026.

[16] Council of Europe. *Report on Semantic Categorisation of Judicial Decisions in Caselaw Databases with Recommendations*. TJENI Project Results, Council of Europe, 2025 disponibile all'indirizzo: https://rm.coe.int/report-on-semantic-categorisation-of-judicial-decisions/1680aeb729.

[17] Myra Cheng et al., Sycophantic AI decreases prosocial intentions and promotes dependence. Science391,eaec8352(2026). DOI:10.1126/science.aec8352.

[18] S. Vagnoni; M. Palmirani, Fostering Deliberative Democracy in the Digital Age: An Ai-Powered Platform for Enhanced Citizen Engagement in Legislative Processes, in: Proceedings of the 2025 Eleventh International Conference on eDemocracy & eGovernment (ICEDEG), 2025, pp. 339 - 341 (atti di: 2025 Eleventh International Conference on eDemocracy & eGovernment (ICEDEG), Bern, Switzerland, 18-20 June 2025).

[19] Kennan, Ariel, Lisa Singh, Alessandra Garcia Guevara, Mohamed Ahmed e Jason Goodman. AI-Powered Rules as Code: Experiments with Public Benefits Policy. Washington, DC: Massive Data Institute, Georgetown University, 2025; Salvi, Francesco, Manoel Horta Ribeiro, Riccardo Gallotti, e Robert West. "On the Conversational Persuasiveness of Large Language Models to Socially Relevant Topics." Proceedings of the National Academy of Sciences (PNAS) 121, n. 24 (Giugno 2024).

[20] Domanti, U., Campidelli, L., Agnoli, S., & De Angeli, A. (2026). Are Semantic Networks Associated with Idea Originality in Artificial Creativity? A Comparison with Human Agents. Accettato a CHI 2026 (ACM Conference on Human Factors in Computing Systems).

Anche il CSM nella sua delibera dell'8 ottobre 2025, approvando le "Raccomandazioni sull'uso dell'intelligenza artificiale", svolge simili considerazioni [21], ma tratteggia anche importanti raccomandazioni che portano a mitigare tali rischi. Anche la Commissione europea per l'Efficienza della Giustizia - CEPEJ traccia linee guida per non cadere nel revisionismo e luddismo, ma nel contempo non sottrarsi alla doverosa analisi dei rischi[22][23].

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 Forum[24] 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 credenze[25]. 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 invisibile[26].

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

---

[21] Consiglio Superiore della Magistratura, Delibera 8 ottobre 2025, Raccomandazioni sull'uso dell'intelligenza artificiale nell'amministrazione della giustizia, https://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."

[22] European Commission for the Efficiency of Justice (CEPEJ). *Guidelines on the Use of Generative Artificial Intelligence for Courts*. CEPEJ(2025)18Final. Strasbourg: Council of Europe, December 2025. https://rm.coe.int/cepej-2025-18final-en-guidelines-on-the-use-of-generative-ai-for-courts/48802a4ad1.

[23] European Commission for the Efficiency of Justice (CEPEJ). *Guidelines for the Online Publication of Judicial Decisions and Access to Legal Knowledge*. CEPEJ(2024)9Rev. Strasbourg: Council of Europe, December 2024. https://rm.coe.int/cepej-2024-9-guidelines-on-the-online-publication-of-judicial-decision/1680b2d0de

[24] World Economic Forum, Global Risks Report 2026 (Geneva: World Economic Forum, 2026), https://reports.weforum.org/docs/WEF_Global_Risks_Report_2026.pdf.

[25] Sterling Williams-Ceci et al., Biased AI writing assistants shift users' attitudes on societal issues. Sci. Adv.12,eadw5578(2026).DOI:10.1126/sciadv.adw5578.

[26] K. Hackenburg et al., The levers of political persuasion with conversational artificial intelligence.Science390,eaea3884(2025). DOI:10.1126/science.aea3884.

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 DSA[27], restano ampie zone grige 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*[28] 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 cosiddetta "personalizzazione dinamica".

Il lavoro pubblicato su *Science*[29], introduce un meccanismo ancora più problematico: il cosiddetto "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 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 profilandolo[30]. 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 ad 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 così detti 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

---

[27] European Union. *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (Digital Services Act)*. Official Journal of the European Union L 277 (27 October 2022): 1–102. https://eur-lex.europa.eu/eli/reg/2022/2065/oj.

[28] Lin, Hause, Gabriela Czarnek, Benjamin Lewis, Joshua P. White, Adam J. Berinsky, Thomas Costello, Gordon Pennycook, et al. Persuading Voters Using Human–Artificial Intelligence Dialogues. Nature 648 (2025): 394–401.

[29] Hackenburg, Kobi, Ben M. Tappin, Luke Hewitt, Ed Saunders, Sid Black, Hause Lin, Catherine Fist, Helen Margetts, David G. Rand, and Christopher Summerfield. *The Levers of Political Persuasion with Conversational AI*. *Science* 390, no. 6777 (2025).

[30] V. Lerman, Y. Dover; A closer look at how large language models 'trust' humans: patterns and biases. *Proc. A* 1 April 2026; 482 (2335): 20251113. https://doi.org/10.1098/.

AI dedicati a questo scopo possono intercettare fallacie logiche negli argomenti degli esseri umani ma anche nelle conversazioni prodotte dall'IA Generativa. 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*[31]: 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*[32]: la normativa è progettata sulla base degli obiettivi di policy e dei suoi effetti.

Questi modelli sono complementari e da usare caso per caso usando una metodologia chiamata IA ibrida[3334], 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 dominante[35].

Nel report HAI2026 di Stanford vi sono benchmarking di modelli LLM effettuati su giurisprudenza USA e canadese[36]: 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 di questi parametri:

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

---

[31] R. Kennedy, *Rules as Code and the Rule of Law: Ensuring Effective Judicial Review of Administration by Software*, *Law, Innovation and Technology* 16, no. 1 (2024): 170–193, https://doi.org/10.1080/17579961.2024.2313801. J. Mohun and A. Roberts, *Cracking the Code: Rulemaking for Humans and Machines*, OECD Working Papers on Public Governance, no. 42 (Paris: OECD Publishing, 2020), https://doi.org/10.1787/3afe6ba5-en

[32] European Commission, DG DIGIT, *A Semantic Approach to Digital-Ready Policymaking: The Legislative Financial and Digital Statement (LFDS)*, presentation at ENDORSE Conference, October 2025, Interoperable Europe Portal.

[33] Palmirani, Monica, Salvatore Sapienza, and Kevin Ashley. *A Hybrid Artificial Intelligence Methodology for Legal Analysis*. *BioLaw Journal – Rivista di BioDiritto*, no. 3 (2024): 389–409. https://doi.org/10.15168/2284-4503-3206.

[34] Palmirani, Monica, Michele Corazza, Generoso Longo, and Salvatore Sapienza. *Hybrid AI to Enhance Legal Drafting with LEOS*. Luxembourg: European Commission, dicembre 2025. https://doi.org/10.2799/1298773.

[35] Goodenough, Oliver R., and Paul J. Carlson. *Words or Code First? Is the Legacy Document or a Code Statement the Better Starting Point for Complexity-Reducing Legal Automation? Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 382, no. 2270 (2024)

[36] https://www.vals.ai/benchmarks/case_law_v2. https://www.vals.ai/benchmarks/legal_bench

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[37]. In questo contesto, gli *agentic AI* offrono ulteriori possibilità[38], permettendo di suddividere compiti complessi tra agenti specializzati e di gestire processi di ragionamento grazie a meccanismi di memoria integrata da regole deterministiche.

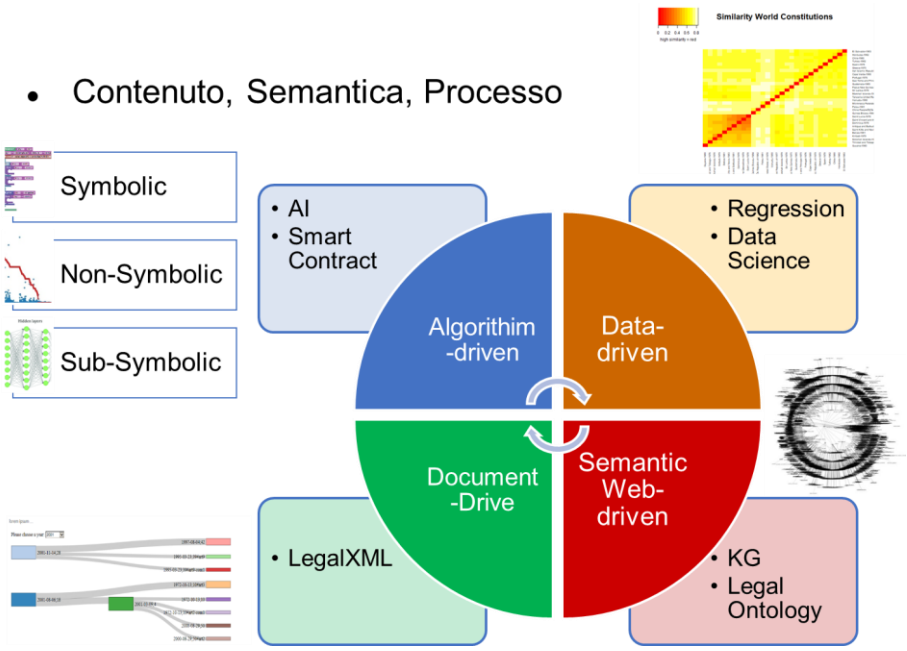

Figura 2 – Impostazione ibrida all'IA integrata con altre tecnologie

I 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, OECD, 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 (si veda anche l'art. 22 del GDPR), 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[39] come capacità di produrre una spiegazione, semmai con una iterazione persona-macchina tale da inquadrare correttamente i parametri della spiegazione attesa. I sistemi di IA applicati

---

[37] OASIS, *LegalDocML (Akoma Ntoso) Version 1.0*, OASIS Standard (Burlington, MA: Organization for the Advancement of Structured Information Standards, 2018), https://docs.oasis-open.org/legaldocml/akn-core/v1.0/akn-core-v1.0.html.

[38] Floridi, Luciano. *La differenza fondamentale. Artificial Agency: una nuova filosofia dell'intelligenza artificiale*. Milano: Mondadori, 2025.

[39] Sapienza, Salvatore, and Monica Palmirani. "Algorithmic Knowability: A Unified Approach to Explanations in the AI Act." In *Explainable Artificial Intelligence (xAI 2025)*. Springer, 2025. Palmirani, Monica, and Salvatore Sapienza. "Big Data, Explanations and Knowability." *Ragion Pratica* 2021: 349–364.

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 controfattuali[40] al fine di evidenziare lacune nella robustezza del ragionamento giuridico esito dell'elaborazione. Le categorie del ragionamento dell'essere umano seguono processi diversi[41] e non si possono paragonare, ma si possono integrare se l'utente finale riesce ad 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[42]), che utilizzano per esempio strumenti di intelligenza artificiale generativa generalistici. La regolazione europea e le linee guida collegate[43] in tema di intelligenza artificiale nell'art. 25[44] sposta 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 ad 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.

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

Tipo di modifica	Potenza di computazione utilizzata	Esito
Prompt tuning, RAG, integrazione	Minimo	Rimane deployer
Fine-tuning leggero (LoRA, adapter)	Solitamente < 1/3 di 10^23 Flops	Rimane deployer
Fine-tuning massiccio o re-training parziale	> 1/3 di 10^23 Flops	Diventa provider GPAI
Fine-tuning massiccio o re-training parziale	>10^25 Flops	Diventa provider GPAI a rischio sistemico

[40] Il pensiero controfattuale è stato evocato dal Senato della Repubblica come strumento per esercitare il pensiero critico nei confronti del dialogo con l'IA.
https://www.senato.it/application/xmanager/projects/leg19/file/repository/UVI/strumenti/Strumenti_1.pdf
[41] Barrett, L.F., Miller, E.K. Categorization is 'baked' into the brain. *Nat. Rev. Neurosci.* (2026). https://doi.org/10.1038/s41583-026-01036-2
[42] Si veda l'articolo 3, Definizioni dell'AI Act, in particolare i punti dal 63 al
[43] European Commission. *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. Brussels, November 19, 2025.
[44] L'Articolo 25 dell'EU AI Act (Regolamento UE 2024/1689) disciplina le "Responsabilità lungo la catena del valore dell'IA".

Cambio scopo → high-risk	Indipendente dalla capacità di calcolo	Diventa provider high risk
Rebranding + immissione sul mercato	In ogni caso	Il deployer diventa **provider**

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'ultima può articolarsi in una pluralità di soggetti, frequentemente qualificati, spesso in modo non esplicito, come "fornitori a valle"[45] o, con terminologia di settore, come integratori di sistemi di intelligenza artificiale. Ricordiamo che l'AI Act elenca fra le applicazioni ad altro rischio anche le ricerche documentali delle banche dati giuridiche (legislative e giurisprudenziali)[46] 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 un caso), incidendo sulla funzione valutativa dei magistrati.

In ogni caso 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. 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[47] 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 costante[48][49] e una policy chiara di design dei sistemi di IA che tenga in conto del contesto e della tradizione giuridica.

---

[45] Vedi art. 3, par. 1, n. 65, Reg. (UE) 2024/1689 (AI Act). Nella versione in lingua inglese il medesimo attore è definito *downstream provider*.

[46] L'allegato III include nel punto 5 "Accesso a servizi privati essenziali e a prestazioni e servizi pubblici essenziali e fruizione degli stessi", e nel punto 8, lettera 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;"

[47] Almasoud, A.S., Idowu, J.A. Algorithmic fairness in predictive policing. *AI Ethics* **5**, 2323–2337 (2025). https://doi.org/10.1007/s43681-024-00541-3.

[48] Francesca Lagioia, Riccardo Rovatti, and Giovanni Sartor, *Algorithmic Fairness through Group Parities? The Case of COMPAS-SAPMOC*, *AI & Society* 38, no. 2 (2023): 459–478.

[49] Cofone, Ignacio, and Poomsit Khern-amnuai. The Overstated Cost of AI Fairness in Criminal Justice. Indiana Law Journal 100, no. 4 (2025).

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 necessitino 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 abbracciano diverse competenze possono raggiungere risultati duraturi che non risentono delle mode del momento. L'entusiasmo deve essere compensato da un bilanciamento fra innovazione e prudenza, nel rispetto dei diritti fondamentali, dei principi costituzionali, delle istituzioni democratiche.