

LINEE GUIDA SULL'USO DELL'INTELLIGENZA ARTIFICIALE GENERATIVA PER I TRIBUNALI

Strasburgo, 19 dicembre 2025

CEPEJ(2025)18Final

**COMMISSIONE EUROPEA PER L'EFFICIENZA DELLA GIUSTIZIA
(CEPEJ)**

Adottate alla 45^a riunione plenaria della CEPEJ (Strasburgo, 4 e 5
dicembre 2025)

Indice

Sintesi esecutiva	
3	
1. Introduzione	
4	
• 1.1. Premessa	4
• 1.2. L'intelligenza artificiale generativa	4
• 1.3. L'IA generativa nel settore della giustizia	5
2. Le Linee Guida	
6	
• 2.1. Aspetti operativi per l'IA generativa nei tribunali	6

○ 2.1.1. Partire dall'individuazione del problema, non dalla soluzione tecnica	6
○ 2.1.2. Individuare le aree idonee all'utilizzo dell'IA generativa	7
○ 2.1.3. Essere consapevoli delle applicazioni non idonee	7
○ 2.1.4. Preferire strumenti personalizzati e sicuri rispetto a soluzioni generiche ..	8
○ 2.1.5. Assicurare una gestione efficace nell'implementazione dei sistemi di IA generativa	10
○ 2.1.6. Essere consapevoli dell'uso dell'IA generativa da parte delle parti	11
○ 2.1.7. Garantire consapevolezza e formazione degli utenti di IA	12
• 2.2. Aspetti normativi per l'IA generativa nei tribunali	13
○ 2.2.1. Diritto a un equo processo e indipendenza giudiziaria (Art. 6 CEDU) ...	13
○ 2.2.2. Diritto alla libertà e alla sicurezza (Art. 5 CEDU)	15
○ 2.2.3. Diritto alla privacy e alla protezione dei dati (Art. 8 CEDU)	15
○ 2.2.4. Diritto a un ricorso effettivo (Art. 13 CEDU)	16
○ 2.2.5. Divieto di discriminazione (Art. 14 CEDU e Protocollo n. 12)	16
○ 2.2.6. Ulteriori garanzie	16

3. GLOSSARIO	17
---------------------------	-----------

SINTESI ESECUTIVA

Le presenti Linee Guida forniscono un quadro di riferimento per l'implementazione dell'intelligenza artificiale (IA) generativa nell'amministrazione della giustizia, con particolare attenzione alla

certezza del diritto, all'indipendenza della magistratura con un'efficace supervisione umana, alla trasparenza e alla tracciabilità.

In primo luogo, gli strumenti generici forniti dalle grandi aziende tecnologiche rappresentano solo un primo passo; è necessario orientarsi verso soluzioni specializzate e, ove appropriato, verso modelli linguistici di grandi dimensioni (Large Language Models - LLM) locali o implementazioni con controllo sovrano sui dati e sull'infrastruttura. È essenziale che gli strumenti siano adattati ai processi giudiziari e implementati sotto controllo pubblico.

L'implementazione di tali soluzioni deve essere conforme alla Convenzione Europea dei Diritti dell'Uomo (CEDU), alla Convenzione Quadro del Consiglio d'Europa sull'Intelligenza Artificiale (CETS 225), alla Convenzione 108+ per la protezione delle persone rispetto al trattamento dei dati personali e alla Carta Etica Europea della CEPEJ sull'Uso dell'Intelligenza Artificiale nei Sistemi Giudiziari e nel loro Ambiente.

Le presenti Linee Guida si fondano sui seguenti principi cardine: l'esercizio del potere giurisdizionale è responsabilità esclusiva dei tribunali; l'accesso effettivo a un giudice persona fisica è sempre garantito; i risultati prodotti dall'IA generativa non sono mai vincolanti; l'utilizzo dell'IA generativa nelle attività giudiziarie e nella redazione delle decisioni deve essere trasparente. I requisiti in materia di privacy impongono la pseudonimizzazione dei dati e il controllo esclusivo dei dati e dell'infrastruttura da parte dell'autorità pubblica competente per la fornitura dei servizi informatici alla magistratura. È fondamentale considerare la legalità, l'indipendenza giudiziaria, il principio di non discriminazione, il diritto a un ricorso effettivo e la tracciabilità e spiegabilità della motivazione.

Si propone un'implementazione graduale, con progetti pilota, parametri di valutazione e la possibilità di tornare sui propri passi, se necessario. Questo approccio si fonda sul principio di sussidiarietà, che enfatizza l'individuazione delle esigenze funzionali e la considerazione prioritaria delle soluzioni tecnologiche convenzionali. L'IA generativa dovrebbe essere implementata solo quando le soluzioni alternative siano ritenute inefficaci o inefficienti.

Il suo utilizzo può essere vantaggioso in diversi contesti professionali, tra cui la gestione dei procedimenti, l'elaborazione dei documenti, i servizi di informazione al pubblico e il supporto alla redazione di atti giudiziari di routine. Quando viene utilizzata per assistere le autorità giudiziarie nella ricerca e interpretazione dei fatti e del diritto, nell'applicazione della legge a situazioni concrete, o in qualsiasi altro supporto all'attività decisionale, devono essere rispettati requisiti rigorosi. Per garantire un'implementazione efficace, è essenziale condurre valutazioni approfondite ex ante dei rischi e degli impatti prima di qualsiasi implementazione. Inoltre, l'implementazione di meccanismi continui di monitoraggio e feedback consentirà sia al tribunale che all'amministrazione della giustizia di apportare modifiche e miglioramenti. Infine, la formazione di tutto il personale giudiziario e degli utenti del sistema è fondamentale per garantire l'uso efficace del sistema.

Infine, lo Stato deve rimanere responsabile per qualsiasi danno derivante dall'uso dei sistemi di IA generativa nell'amministrazione della giustizia, assicurando che gli utenti mantengano vie di ricorso e rimedi effettivi in caso di pregiudizio.

1. INTRODUZIONE

1.1. Premessa

1. Le presenti linee guida mirano a fornire indicazioni pratiche su come implementare in modo sicuro l'intelligenza artificiale (IA) generativa nell'amministrazione della giustizia. Sono rivolte alle autorità responsabili dell'amministrazione della giustizia e ai professionisti del settore giudiziario. Dovrebbero essere periodicamente riesaminate e aggiornate in linea con i recenti sviluppi tecnologici.
2. L'uso dell'IA nella giustizia deve rispettare i diritti garantiti dalla Convenzione Europea dei Diritti dell'Uomo (CEDU), dalla Convenzione Quadro del Consiglio d'Europa sull'Intelligenza Artificiale e i Diritti Umani, la Democrazia e lo Stato di Diritto (CETS225), la protezione dei dati personali alla luce della Convenzione del Consiglio d'Europa per la Protezione dei Dati Personali (CETS108), nonché i principi stabiliti nella Carta Etica

Europea sull'Uso dell'Intelligenza Artificiale nei Sistemi Giudiziari e nel loro Ambiente (Carta CEPEJ), e ove applicabile il Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024 che stabilisce norme armonizzate sull'intelligenza artificiale (AI Act).

3. Le presenti linee guida sono state elaborate dal Gruppo di Lavoro CEPEJ sulla Cybergiustizia e l'IA (CEPEJ-GT-CYBERJUST) sulla base del contributo di esperti[1]. Esse tengono conto del lavoro preliminare della CEPEJ, come la "Nota informativa sull'uso dell'IA generativa da parte dei professionisti giudiziari in un contesto lavorativo", le "Riflessioni del Comitato Consultivo sull'Intelligenza Artificiale della CEPEJ su come far funzionare l'intelligenza artificiale per la magistratura" e il "1° Rapporto sull'uso dell'intelligenza artificiale (IA) nella magistratura, basato sulle informazioni contenute nel Centro di Risorse della CEPEJ sulla Cybergiustizia e l'IA" e sono illustrate attraverso strumenti presentati nel Centro di Risorse della CEPEJ sulla Cybergiustizia e l'IA.

1.2. L'intelligenza artificiale generativa

4. L'IA generativa è un ramo dell'IA che si concentra sull'insegnare alle macchine a creare nuovi contenuti che imitano in modo convincente le produzioni umane. Invece di limitarsi a classificare o prevedere i dati, questi sistemi apprendono da grandi insiemi di dati per identificare modelli e strutture sottostanti, consentendo loro di generare testi, immagini, musica, video e altri tipi di contenuti[2].
5. Nel campo della generazione di testi, l'evoluzione degli strumenti di "elaborazione del linguaggio naturale" ha permesso ai software automatizzati di interpretare e generare correttamente il linguaggio in modo naturale. In precedenza, la comprensione del linguaggio umano rappresentava una sfida per i software di IA a causa della sua ambiguità e complessità, ma con i progressi nell'apprendimento profondo, i modelli sono diventati capaci di apprendere i modelli del nostro linguaggio e produrre risposte o testi coerenti. Ciò ha comportato un cambio di paradigma nel modo in cui le macchine elaborano e generano contenuti testuali.

6. La tecnologia alla base dell'IA generativa si basa sull'apprendimento profondo e su architetture avanzate di reti neurali. In sostanza, questi modelli, noti come Large Language Models o LLM, sono addestrati con volumi di dati vastissimi per identificare modelli complessi (che si tratti di sequenze di parole, pixel di immagini o note musicali) che consentono di generare nuovi contenuti che imitano le caratteristiche del materiale originale.
7. Nonostante la sua efficacia in un'ampia gamma di compiti e l'adozione nei chatbot, l'IA generativa presenta attualmente delle limitazioni, tra cui le "allucinazioni". Nella generazione del linguaggio naturale, un'allucinazione è un output che il modello produce senza un supporto fedele nell'input o nella conoscenza di riferimento: può contraddire esplicitamente il testo di origine (intrinseca) o aggiungere dati impossibili o falsi (estrinseca). In sostanza, si tratta di informazioni inventate che il sistema presenta come se fossero fatti[3].
8. Altre limitazioni degne di nota riguardano potenziali problematiche relative ai diritti d'autore derivanti dai dati di addestramento originali di questi modelli, nonché la potenziale divulgazione non autorizzata dei dati in input a terzi e il loro riutilizzo per addestrare il modello di IA generativa, potenzialmente in violazione della normativa applicabile sulla protezione dei dati.

1.3. L'IA generativa nel settore della giustizia

9. L'uso di soluzioni di IA generativa da parte dei professionisti della giustizia è in crescita, sia attraverso la generazione di riassunti, la correzione di testi, la redazione di fondamenti giuridici basati sulla giurisprudenza, l'interazione con i soggetti processuali (spiegazione delle sentenze ai clienti), ecc. Secondo il rapporto Future Ready Lawyer 2024[4], che ha intervistato 712 professionisti legali negli Stati Uniti e in nove paesi europei, il 68% degli studi legali e il 76% degli uffici legali aziendali utilizzano l'IA generativa almeno settimanalmente, con oltre un terzo che la utilizza quotidianamente (35% negli uffici legali aziendali e 33% negli studi legali). In Europa, i Paesi Bassi guidano l'uso quotidiano dell'IA generativa con il 42%, seguiti dalla Germania con il 38%. L'uso principale di questi strumenti

- riguarda l'automazione dei documenti (67%), la gestione del tempo (66%) e la pianificazione strategica (62%). Inoltre, il 58% degli studi legali prevede di aumentare i propri investimenti in IA nei prossimi tre anni, e il 73% degli uffici legali aziendali.
10. Secondo un sondaggio tra i membri della Rete Europea di Cybergiustizia della CEPEJ dell'ottobre 2025, il 46% degli intervistati[5] ha confermato l'uso dell'IA generativa nei tribunali, di cui il 74% si riferisce ad applicazioni standard/pronte all'uso (ad es. Microsoft Copilot, Chat GPT, DeepL) e il 52% ad applicazioni personalizzate/su misura (entrambe le opzioni possibili). Gli usi più menzionati dell'IA generativa riguardano la sintesi di documenti e altre funzioni di supporto editoriale. Nessun intervistato ha menzionato l'uso dell'IA generativa per la stima quantitativa degli esiti legali (vedi Figura 1).
 11. Diversi paesi stanno facilitando l'uso di strumenti di IA generativa nella pubblica amministrazione e nei tribunali attraverso accordi di riservatezza, protezione dei dati di utilizzo e garanzie che tali dati non saranno riutilizzati per il riaddestramento dei modelli[6][7].
 12. Nel settore legale, l'uso dell'IA generativa ha portato a cause legali con citazioni giuridiche false[8][9], che a volte hanno portato a sanzioni per i suoi utenti[10]. Poiché le "allucinazioni" sono inevitabili[11] o nella migliore delle ipotesi mitigabili[12], secondo lo stato dell'arte attuale, questo è un rischio da considerare quando si impiega l'IA generativa nella giustizia, senza adeguate garanzie di supervisione umana.
-

2. LE LINEE GUIDA

2.1. Aspetti operativi per l'IA generativa nei tribunali

2.1.1. Partire dall'individuazione del problema, non dalla soluzione tecnica

13. Il principio di neutralità tecnologica stabilisce che le pubbliche amministrazioni dovrebbero concentrarsi sulle esigenze funzionali e, su questa base, selezionare soluzioni tecnologiche che possano essere adattate nel tempo. Questo approccio mira a ridurre al minimo le dipendenze tecnologiche, evitare di imporre

- implementazioni tecniche o prodotti specifici e rimanere adattabile in un ambiente tecnologico in rapida evoluzione. Questo principio dovrebbe estendersi all'uso dell'IA generativa all'interno dell'amministrazione giudiziaria attraverso un principio di sussidiarietà, per cui le amministrazioni giudiziarie identificano in primo luogo chiaramente le esigenze funzionali e in secondo luogo considerano diverse soluzioni tecnologiche, convenzionali o innovative, prima di optare per la più adatta e meno rischiosa. Ciò non dovrebbe impedire ai tribunali di impegnarsi continuamente per migliorare l'efficienza della giustizia con il supporto della tecnologia, ove appropriato, né di astenersi dall'introdurre l'IA generativa in casi d'uso ben consolidati che soddisfano esigenze funzionali ben identificate.
14. Sia i tribunali che le amministrazioni della giustizia affrontano varie sfide, tra cui carenza di personale, alti volumi di cause di massa, compiti di redazione ripetitivi, gestione inefficiente dei fascicoli, ritardi procedurali, difficoltà nel reperire la giurisprudenza, problemi di comunicazione con gli utenti del sistema giudiziario e mancanza di interoperabilità tra i sistemi informatici attualmente in uso.
 15. Una volta identificate le esigenze operative e i problemi all'interno dell'amministrazione, il passo successivo è la selezione di soluzioni tecniche appropriate in conformità con i principi di neutralità tecnologica e sussidiarietà dell'IA.
 16. Nella selezione delle soluzioni tecniche più appropriate, dovrebbe essere considerata l'efficienza in termini di costi. Ciò dovrebbe includere fattori quali il consumo energetico, le misure di sicurezza informatica, le licenze, i costi di manutenzione, sviluppo e consulenza, oltre al prezzo di acquisto iniziale.

2.1.2. Individuare le aree idonee all'utilizzo dell'IA generativa[13]

2.1.2.1. Gestione dei procedimenti e amministrazione giudiziaria

17. L'IA generativa potrebbe eseguire la gestione predittiva del carico di lavoro analizzando i modelli storici e creare archivi di giurisprudenza organizzati per tipo di causa e materia. Questi strumenti potrebbero facilitare l'identificazione di modelli argomentativi in rivendicazioni simili, consentendo l'individuazione di procedimenti che possono essere risolti in modo uniforme quando sollevano la stessa questione giuridica,

o aiutare nell'assegnazione dei casi. L'IA generativa potrebbe anche consentire il riconoscimento automatico delle questioni controverse confrontando i reclami e le risposte.

2.1.2.2. Gestione documentale ed elaborazione delle informazioni

18. L'IA generativa potrebbe generare automaticamente verbali di udienza, redigere indici tematici e cronologici per fascicoli voluminosi ed estrarre dati rilevanti da documenti non strutturati. Allo stesso modo, potrebbe essere utilizzata per ottenere informazioni chiave, come nomi, date e importi, incorporandole automaticamente in modelli predefiniti.

2.1.2.3. Servizi informativi e assistenza agli utenti

19. L'IA generativa potrebbe consentire la creazione di chatbot specializzati per materia giuridica e sistemi di orientamento procedurale conversazionale per i cittadini. Queste applicazioni devono essere progettate con particolare attenzione all'accuratezza delle informazioni fornite e alla chiarezza della comunicazione con utenti non specializzati, impiegando un linguaggio semplice.

2.1.2.4. Supporto alla redazione e documentazione giudiziaria

20. L'IA generativa potrebbe essere utilizzata per la creazione automatica di atti procedurali standard, per l'assistenza nella redazione di ordinanze e per la verifica della coerenza dei documenti. Potrebbe generare riassunti strutturati di sentenze, atti procedurali e ricorsi, richiedendo la verifica giudiziaria che il riassunto corrisponda fedelmente all'originale. Potrebbe anche facilitare la preparazione di documenti, come relazioni, schemi o mappe mentali, che presentano una panoramica dei punti di accordo e disaccordo.

21. Inoltre, potrebbe essere utilizzata per l'analisi di atti introduttivi e difese con suggerimenti di giurisprudenza pertinente; identificazione e sintesi della giurisprudenza rilevante per ciascun caso; analisi preliminare dei ricorsi in merito al rispetto dei requisiti formali e tecnici; e collegamento degli argomenti tra diversi scritti processuali, producendo documenti organizzati in cui la risposta del convenuto è sistematicamente associata a ciascuno degli argomenti dell'attore.

2.1.3. Essere consapevoli delle applicazioni non idonee

22. L'AI Act dell'UE (Regolamento (UE) 2024/1689) stabilisce divieti specifici sulle pratiche di IA particolarmente rilevanti per l'amministrazione della giustizia e potenzialmente per l'IA generativa. L'articolo 5 dell'Atto stabilisce le seguenti pratiche di IA vietate, che possono servire da esempio anche per i paesi al di fuori del suo ambito di applicazione: sistemi di IA che impiegano tecniche subliminali o manipolative (Art. 5(1)(a)); sistemi di IA che sfruttano vulnerabilità (Art. 5(1)(b)); sistemi di valutazione sociale (Art. 5(1)(c)); sistemi di valutazione del rischio per prevedere reati (Art. 5(1)(d)); sistemi di espansione di database di riconoscimento facciale (Art. 5(1)(e)); sistemi di inferenza delle emozioni (Art. 5(1)(f)).
23. Inoltre, sebbene non esplicitamente vietati, l'AI Act dell'UE classifica alcuni sistemi di IA nel settore giudiziario come ad alto rischio ai sensi dell'Allegato III, in particolare quelli destinati ad assistere le autorità giudiziarie nella ricerca e interpretazione di fatti e diritto, e nell'applicazione del diritto a situazioni concrete. Questi sistemi richiedono il rispetto di requisiti rigorosi ma non sono vietati in assoluto.
24. Sebbene gli LLM possano essere utili nella raccolta delle prove nella fase investigativa, non dovrebbero sostituire la valutazione giudiziale delle prove, una volta introdotte nel processo. Sebbene esistano filoni di IA logica e neurosimbolica che integrano modelli causali[14], l'IA generativa manca della capacità di "teorizzazione prospettica e sviluppo di logica causale"[15][16].
25. Il processo decisionale dovrebbe rimanere sotto controllo umano. I casi che richiedono una nuova interpretazione normativa o l'analisi di principi giuridici astratti non dovrebbero essere affidati all'IA, dato che essa manca di "comprensione intrinseca di ciò che è vero o falso nella realtà"[17]. Allo stesso modo, l'uso dell'IA generativa dovrebbe essere evitato in materie che richiedono l'interpretazione evolutiva di normative recenti, dato che l'IA generativa può mostrare limitazioni nell'adattare la propria comprensione a regolamenti innovativi se non è stata specificamente addestrata con essi[18].
26. L'analisi della costituzionalità o della conformità ai trattati internazionali richiede un livello di interpretazione assiologica

che supera le attuali capacità dell'IA. Anche se ciò non fosse vero in futuro, la questione può essere lasciata solo al giudizio umano, poiché riguarda il modo in cui i cittadini hanno deciso di governarsi. Pertanto, lo "zeitgeist" di ogni momento influenzerà il "controllo costituzionale e di convenzionalità", e questo non può essere lasciato alla discrezione di un'IA.

2.1.4. Preferire strumenti personalizzati e sicuri rispetto a soluzioni generiche pronte all'uso

27. La distinzione tra strumenti commerciali generici di IA generativa e IA generativa giudiziaria personalizzata è altamente rilevante per l'amministrazione della giustizia perché comporta non solo specifiche tecniche ma questioni fondamentali di protezione dei dati, indipendenza giudiziaria e integrità dei procedimenti legali.
28. Gli strumenti commerciali gratuiti di IA generativa presentano rischi significativi. Quando i professionisti giudiziari utilizzano servizi di IA generativa non ufficiali, siano essi soluzioni commerciali gratuite o a pagamento, rischiano inavvertitamente di incorrere in un trasferimento internazionale di dati, principalmente verso data center negli Stati Uniti, ospitati dalle società che forniscono questi servizi. Questi dati diventano successivamente disponibili per scopi di riaddestramento del modello, rimuovendoli efficacemente dal controllo amministrativo giudiziario. Tali pratiche violano i principi fondamentali di protezione dei dati stabiliti nella Convenzione 108+[19] e creano vulnerabilità inaccettabili nell'ecosistema informativo giudiziario.
29. Le soluzioni commerciali generiche di IA generativa, anche quando offerte tramite abbonamenti a pagamento, tipicamente mancano della formazione specializzata necessaria per il ragionamento giuridico e la documentazione giudiziaria. Questi strumenti, progettati per l'uso generale dei consumatori, dimostrano una comprensione limitata della terminologia giuridica, dei requisiti procedurali e dei modelli di ragionamento sfumati caratteristici del lavoro giudiziario. L'assenza di integrazione con database legali verificati aumenta il rischio di "allucinazioni", la generazione di disposizioni legali inesistenti o

giurisprudenza falsamente rappresentata, che pone gravi minacce alla certezza del diritto.

30. In caso di ricorso a soluzioni commerciali, l'implementazione dell'IA generativa all'interno dell'amministrazione della giustizia dovrebbe essere mediata esclusivamente attraverso accordi contrattuali formali tra autorità pubbliche e fornitori di servizi di IA. In caso di ricorso a soluzioni interne, l'implementazione dell'IA generativa all'interno dell'amministrazione della giustizia dovrebbe essere mediata attraverso una politica sull'uso dell'IA nell'amministrazione della giustizia[20] o uno schema più ampio di sicurezza delle informazioni[21]. Questi contratti o politiche dovrebbero stabilire quadri completi che affrontino:

- Protocolli rigorosi di governance dei dati che garantiscano l'assenza di trasferimenti non autorizzati a terzi
- Procedure obbligatorie di distruzione dei dati a seguito di periodi di conservazione specificati
- Divieti espliciti contro l'uso dei dati giudiziari per il riaddestramento del modello senza espressa autorizzazione
- Chiara delimitazione della responsabilità per violazioni dei dati o guasti del sistema
- Conformità con le normative nazionali ed europee sulla protezione dei dati
- Requisiti di audit di sicurezza regolari e test di penetrazione
- Procedure trasparenti di risposta agli incidenti
- Meccanismi per il coinvolgimento continuo degli stakeholder (ad es. consultazioni pubbliche, comitati di supervisione, valutazioni periodiche)

31. I modelli specializzati di IA generativa giudiziaria rappresentano un requisito critico per un'implementazione efficace dell'IA in contesti legali. Il linguaggio giuridico costituisce un dominio tecnico con terminologia specializzata che spesso diverge significativamente dall'uso comune. Termini come "corrispettivo", "pregiudizio", "ordine pubblico" o "buona fede" assumono significati giuridici specifici che differiscono sostanzialmente dalle loro definizioni quotidiane. Questa complessità semantica richiede modelli di IA generativa addestrati specificamente su corpora legali di ciascuna

giurisdizione, garantendo l'uso accurato della terminologia giuridica tecnica e la corretta interpretazione contestuale.

32. Le soluzioni personalizzate di IA generativa giudiziaria possono incorporare:

- Integrazione con database legali ufficiali e fonti giurisprudenziali verificate
- Addestramento specifico per giurisdizione su testi legali, decisioni giudiziarie e documenti procedurali
- Capacità multilingue che riflettono la diversità linguistica dei sistemi giudiziari europei
- Moduli specializzati per diverse aree del diritto (civile, penale, amministrativo, costituzionale)
- Sistemi integrati di verifica delle citazioni con riferimenti incrociati a registri legali ufficiali
- Rilevamento automatico e segnalazione di potenziali incongruenze legali o anacronismi

33. La sovranità infrastrutturale costituisce una considerazione essenziale. I sistemi giudiziari di IA generativa dovrebbero operare su infrastrutture sotto il controllo esclusivo delle autorità giudiziarie o delle entità governative quando si opta per modelli open-source in esecuzione su server pubblici, o in collaborazione con società fornitrici quando si opta per modelli commerciali.

2.1.5. Assicurare una gestione efficace nell'implementazione dei sistemi di IA generativa

34. L'implementazione dei sistemi di IA generativa nell'amministrazione giudiziaria dovrebbe seguire un approccio completo che include diversi elementi chiave: una strategia di implementazione incrementale con processi appropriati di gestione del cambiamento e fasi pilota con valutazioni per la revisione, attenta considerazione dei modelli di licenza (tecnologia proprietaria rispetto a open-source), requisiti di supporto e formazione continui da parte dei fornitori, audit regolari, monitoraggio e revisione delle prestazioni del sistema di IA e dei rischi associati.
35. Per garantire una migliore comprensione della tecnologia, nonché per ridurre le potenziali barriere all'adozione e facilitare

la gestione del cambiamento, si suggerisce che l'implementazione dell'IA generativa nell'amministrazione della giustizia segua un approccio graduale, passo dopo passo, differenziato in base alla complessità e alla natura delle questioni affrontate:

Fase preparatoria: Prima di qualsiasi implementazione, dovrebbero essere condotti audit di preparazione istituzionale, dovrebbe essere fornita formazione di base al personale, dovrebbero essere stabiliti protocolli di sicurezza e dovrebbero essere creati comitati di supervisione.

Prima fase: Limitata alle funzioni amministrative e di supporto di base che non hanno un impatto diretto sul processo decisionale giudiziario. Questa fase dovrebbe concentrarsi sulla gestione dei documenti e sull'elaborazione delle informazioni, come l'estrazione di dati rilevanti da documenti non strutturati, l'ottenimento di informazioni chiave (nomi, date, importi) e la loro incorporazione automatica in modelli predefiniti. Allo stesso modo, potrebbe includere la gestione del carico di lavoro di base attraverso l'analisi dei modelli storici e l'assistenza nell'assegnazione dei casi. Possono anche essere implementati servizi di informazione di base e assistenza agli utenti, come sistemi di orientamento procedurale in linguaggio semplice per gli utenti e chatbot specializzati per domande procedurali semplici. Questi dovrebbero essere accompagnati da un'avvertenza sui rischi coinvolti. La trascrizione non simultanea delle udienze potrebbe anche essere considerata in questa fase. Sia i tribunali che l'amministrazione della giustizia dovrebbero valutare l'impatto sul carico di lavoro, monitorare la soddisfazione degli utenti e misurare il tempo di adattamento.

Seconda fase: Espansione a funzioni di supporto analitico e di redazione più sofisticate. Questa fase dovrebbe comprendere capacità avanzate di gestione dei documenti e di elaborazione delle informazioni, inclusa la generazione automatica di verbali di udienza e il collegamento degli argomenti tra diversi scritti processuali. Questa fase potrebbe anche estendersi al supporto alla redazione e alla documentazione giudiziaria, con la creazione automatica di atti procedurali standard, assistenza nella redazione di ordinanze di routine e verifica della coerenza dei documenti. Allo stesso modo, con

la creazione di archivi di giurisprudenza organizzati per tipo di causa e materia e il riconoscimento automatico delle questioni controverse confrontando atti introduttivi e difese, nonché la redazione di indici tematici e cronologici per fascicoli voluminosi. Dovrebbe essere introdotta l'analisi della giurisprudenza e il supporto decisionale per l'analisi preliminare dei ricorsi in merito al rispetto dei requisiti formali e tecnici. I tribunali dovrebbero condurre test A/B per il confronto dell'efficienza, valutare il bias decisionale e analizzare i modelli di utilizzo.

Terza fase: Introduzione di funzioni analitiche e consultive più avanzate, mantenendo una rigorosa supervisione giudiziaria. Questa fase richiede una valutazione positiva delle fasi precedenti, l'approvazione specifica per ciascuna sottofase e protocolli di rollback chiaramente definiti. Questa fase può essere ulteriormente suddivisa in tre sottofasi:

- **Sottofase iniziale:** Focus sulla ricerca giurisprudenziale e sull'organizzazione dei documenti attraverso l'analisi di atti introduttivi e difese con suggerimenti di giurisprudenza pertinente, e identificazione e organizzazione sistematica della giurisprudenza rilevante per ciascun caso per la considerazione giudiziaria. Il supporto alla redazione dovrebbe includere la generazione di riassunti strutturati di sentenze, atti procedurali e ricorsi, richiedendo una verifica giudiziaria completa. Le funzioni avanzate di gestione dei casi potrebbero facilitare l'identificazione di modelli argomentativi in rivendicazioni simili e l'individuazione di procedimenti che possono sollevare questioni giuridiche simili.
- **Sottofase intermedia:** Ricerca giurisprudenziale e organizzazione dei documenti potenziate per supportare la preparazione giudiziaria, fornendo precedenti legali sistematicamente organizzati e giurisprudenza rilevante per la considerazione giudiziaria. Il supporto alla redazione potrebbe includere la preparazione di documenti preparatori come relazioni, schemi e mappe mentali che presentano una panoramica dei punti di accordo e disaccordo tra le parti.
- **Sottofase avanzata:** In procedure altamente standardizzate con precedenti legali ben consolidati, l'IA generativa potrebbe assistere fornendo analisi giurisprudenziale completa con

riconoscimento di pattern tra casi simili, identificando precedenti procedurali e potenziali incongruenze per la revisione giudiziaria, e proponendo bozze che incorporano la giurisprudenza rilevante. Allo stesso modo, potrebbe aiutare a redigere modelli strutturati per atti procedurali di routine basati su pratiche giudiziarie consolidate, mentre l'analisi umana obbligatoria e il processo decisionale rimarrebbero in vigore per tutte le determinazioni legali e giudiziarie e l'approvazione finale di ogni output.

36. Tutto quanto sopra dovrebbe essere applicato come progetti pilota in procedimenti di ingiunzione di pagamento, rivendicazioni valutate per importi determinati, procedimenti verbali di piccoli importi, sanzioni amministrative standardizzate o rivendicazioni lavorative per importi non controversi, in cui tutte le prove sono documentali.
37. Le questioni che riguardano i diritti fondamentali, che coinvolgono minori o persone incapaci, o che comportano misure privative della libertà richiedono l'intervento giudiziario umano completo, limitando l'IA generativa a funzioni di supporto tecnico.

2.1.6. Essere consapevoli dell'uso dell'IA generativa da parte delle parti

38. Sia i tribunali che l'amministrazione della giustizia dovrebbero stabilire protocolli chiari per il monitoraggio e la gestione dell'uso dell'IA generativa da parte dei professionisti legali e delle parti nei procedimenti. Ciò può includere la richiesta alle parti e ai loro rappresentanti legali di dichiarare quando sono stati utilizzati sistemi di IA generativa nella preparazione di atti legali, nella ricerca giuridica o nella compilazione delle prove, quando ritenuto appropriato.
39. Le norme procedurali dovrebbero specificare l'obbligo di notificare l'uso dell'IA generativa nella preparazione del caso, con sanzioni appropriate per la mancata divulgazione o l'eccessivo affidamento su contenuti generati dall'IA senza un'adeguata verifica.
40. I tribunali dovrebbero essere particolarmente vigili riguardo a molteplici forme di cattiva condotta legata all'IA, tra cui: la

presentazione di prove fabbricate, citazioni legali inesistenti o contenuti generati dall'IA che non sono stati adeguatamente esaminati da professionisti qualificati; tentativi di manipolare i sistemi di IA attraverso prompt nascosti o istruzioni incorporate nei documenti, come testo invisibile, formattazione insolita o istruzioni microscopiche incorporate progettate per compromettere l'analisi assistita dall'IA; e l'uso dell'IA generativa per creare video fraudolenti o prove grafiche. Tutte queste pratiche devono essere regolamentate sia proceduralmente che penalmente.

41. Per mantenere l'equità procedurale e la parità delle armi, i tribunali dovrebbero garantire che tutte le parti abbiano pari accesso alle informazioni sull'uso dell'IA generativa nei procedimenti e che nessuna parte ottenga un vantaggio ingiusto attraverso l'assistenza dell'IA non divulgata.

2.1.7. Garantire consapevolezza e formazione degli utenti di IA

42. Giudici, pubblici ministeri, personale amministrativo e tutto il personale giudiziario dovrebbero ricevere un'adeguata formazione per comprendere, gestire e valutare criticamente gli strumenti di IA generativa a loro disposizione e quelli sul mercato. Una formazione specializzata e una maggiore consapevolezza delle potenzialità e dei limiti di questi strumenti garantiranno un uso etico ed efficace della tecnologia, preservando sempre il giudizio umano nel processo decisionale.
43. L'implementazione dell'IA nell'amministrazione della giustizia dovrebbe essere caratterizzata da trasparenza e automazione. Per garantire certezza e coerenza, i giudici non dovrebbero dover eseguire operazioni specifiche per attivare il sistema; il suo uso dovrebbe essere predeterminato dalle leggi procedurali.
44. Le competenze chiave dei giudici dovrebbero concentrarsi sul mantenimento e sull'applicazione del loro giudizio critico, una qualità intrinseca e attesa dalla funzione giudiziaria, valutando rigorosamente ogni risultato fornito dall'IA. Ciò comporta evitare un atteggiamento compiacente o passivo verso i suggerimenti tecnologici e adottare invece un atteggiamento critico, persino dissenziente, verso alcune alternative o proposte generate da questi sistemi.

45. Se i giudici devono essere dotati di sistemi di IA generativa con cui possono interagire direttamente, dovrebbero ricevere una formazione specializzata nel loro uso (prompt engineering, context engineering, ecc.). Dovrebbero acquisire una solida comprensione tecnica dei limiti intrinseci di questi modelli: la portata delle loro capacità, le aree in cui tendono a fallire, i loro punti di forza e i tipi di query che non sono tecnicamente fattibili o appropriati.
46. I giudici dovrebbero essere in grado di analizzare e mettere in discussione gli output dell'IA generativa, identificando potenziali bias, incongruenze o limitazioni nel ragionamento. Dovrebbero sviluppare criteri per valutare quando e come è appropriato utilizzare l'IA generativa in diversi tipi di casi e acquisire la capacità di muoversi efficacemente tra fonti tradizionali (come database giurisprudenziali e articoli dottrinali) e sistemi di IA generativa per ottenere le informazioni più rilevanti e affidabili.
47. I giudici dovrebbero essere formati per evitare il cosiddetto "effetto bolla", che si verifica quando un giudice, impressionato dall'apparente efficienza e accuratezza dei sistemi di IA generativa, sviluppa una fiducia eccessiva nella giurisprudenza, nelle relazioni o nei riassunti che forniscono, rischiando l'abbandono della consultazione indipendente dei database giurisprudenziali originali e il disprezzo per i documenti sottostanti.
48. Dovrebbe essere sviluppata una politica di formazione continua per i nuovi strumenti di IA generativa man mano che emergono, nonché per i progressi che l'IA stessa compie in diversi campi che riguardano la giustizia (generazione di immagini, video, voci, ecc.). Le istituzioni giudiziarie dovrebbero stabilire meccanismi sistematici per mantenere i professionisti giudiziari aggiornati sulle capacità emergenti dell'IA e sulle loro potenziali implicazioni per l'amministrazione della giustizia. Ciò dovrebbe includere la comprensione di nuove forme di prove generate dall'IA, il rilevamento di deepfake e contenuti sintetici, e l'adattamento a metodi in evoluzione di manipolazione digitale che possono influenzare i procedimenti giudiziari. Tale formazione dovrebbe essere regolarmente aggiornata per riflettere lo stato dell'arte nella tecnologia di IA generativa e le sue applicazioni in contesti legali. Per garantire la coerenza

nell'applicazione, potrebbero essere considerati curricula di formazione standardizzati.

2.2. Aspetti normativi per l'IA generativa nei tribunali

49. L'implementazione dei sistemi di IA generativa in ambito giudiziario richiede una valutazione approfondita rispetto ai principi giuridici fondamentali stabiliti dai quadri giuridici europei e dalle disposizioni costituzionali nazionali. Per salvaguardare il rispetto di questi principi giuridici, potrebbero essere utilizzati lo Strumento di Valutazione della CEPEJ[22] e la metodologia di analisi d'impatto HUDERIA[23].

2.2.1. Diritto a un equo processo e indipendenza giudiziaria (Art. 6 CEDU)[24]

2.2.1.1. Garanzie del giusto processo

50. I sistemi di IA generativa possono compromettere l'equità procedurale attraverso processi decisionali opachi, raccomandazioni automatizzate non spiegabili o vantaggi procedurali ineguali derivanti da accesso differenziale a strumenti tecnologici avanzati tra le parti[25].

51. Le garanzie tecniche derivanti da queste disposizioni dovrebbero includere l'implementazione di sistemi di IA spiegabili[26], piena tracciabilità dei processi di ragionamento dell'IA[27], pari accesso per tutte le parti agli output generati dall'IA e meccanismi tecnici che garantiscano una reale parità delle armi[28]. I sistemi di IA a supporto delle decisioni giudiziarie dovrebbero fornire output che siano sia dimostrabilmente spiegabili e affidabili, sia comprensibili per esperti non tecnici. Dovrebbero anche implementare misure per ridurre al minimo le imprecisioni, garantire che venga seguito un corretto ragionamento giuridico e proteggere la sicurezza dei dati.

2.2.1.2. Principio di legalità, certezza del diritto e prevedibilità

52. I sistemi di IA generativa che producono "allucinazioni giuridiche" o fanno riferimento a disposizioni legali inesistenti pongono gravi rischi per la certezza del diritto e lo stato di diritto.

53. L'implementazione dell'IA non dovrebbe compromettere la prevedibilità e la coerenza delle decisioni giudiziarie. Le fonti di informazione, i criteri e le metodologie utilizzate dai modelli di IA generativa dovrebbero essere accuratamente documentati e chiaramente giustificati per prevenire l'introduzione di incertezza o arbitrarietà nei processi giudiziari.
54. Le garanzie tecniche dovrebbero includere la validazione automatica delle citazioni legali, l'uso esclusivo di fonti legali ufficiali certificate e sistemi di allerta per contenuti generati che non possono essere verificati rispetto a database legali autorevoli[29].

2.2.1.3. Esclusività dell'autorità giudiziaria

55. Il potere giurisdizionale spetta esclusivamente ai giudici in conformità con le disposizioni costituzionali e legali. Il processo decisionale, il ragionamento giuridico e la valutazione probatoria non dovrebbero essere delegati ai sistemi di IA generativa; tali sistemi possono essere impiegati esclusivamente per compiti ausiliari o preparatori che supportano, ma non sostituiscono mai, il giudizio umano dell'organo giudiziario.

2.2.1.4. Responsabilità e supervisione umana obbligatorie

56. I giudici mantengono la responsabilità ultima ed esclusiva per tutte le decisioni giudiziarie, indipendentemente da qualsiasi assistenza tecnologica impiegata nel processo decisionale. I funzionari giudiziari dovrebbero comprendere pienamente i limiti di qualsiasi sistema di IA generativa utilizzato e dovrebbero esaminare accuratamente tutti gli output generati dall'IA prima di incorporarli in atti o decisioni giudiziarie.

2.2.1.5. Indipendenza giudiziaria e natura non vincolante dell'IA

57. L'indipendenza giudiziaria[30] può essere compromessa quando i giudici subiscono pressioni dirette o indirette per seguire raccomandazioni generate dall'IA, vincolando così la loro libertà di giudizio, interpretazione giuridica e processi decisionali autonomi. Questa indipendenza dovrebbe estendersi alla libertà dalla coercizione tecnologica o dal bias sistematico introdotto attraverso raccomandazioni dell'IA. I giudici dovrebbero mantenere il diritto di rifiutare di ricevere suggerimenti dall'IA.

58. Le raccomandazioni, i suggerimenti o gli output generati dai sistemi di IA non devono mai essere vincolanti per i giudici, né il loro uso deve essere obbligatorio, preservando così pienamente l'indipendenza giudiziaria nell'interpretazione giuridica, nel ragionamento e nel processo decisionale. Ai giudici non deve essere richiesto di giustificare specificamente la loro divergenza da proposte o raccomandazioni generate dai sistemi di IA.
59. La progettazione dei modelli di IA generativa non dovrebbe dare priorità alla massimizzazione delle loro capacità persuasive al fine di ridurre al minimo il rischio che un'indebita influenza psicologica e il bias di automazione abbiano un effetto negativo sull'imparzialità giudiziaria e sulla libertà nella valutazione delle opinioni giudiziarie generate dall'IA[31].
60. Le garanzie tecniche dovrebbero essere proporzionate ai rischi del sistema e includere la progettazione di modelli di IA generativa rigorosamente come strumenti ausiliari non vincolanti, l'eliminazione di qualsiasi meccanismo che obblighi o prema i giudici a giustificare il discostarsi dai suggerimenti algoritmici, l'esecuzione di valutazioni rigorose prima di qualsiasi uso implementato di strumenti di IA, la conduzione di audit tecnici indipendenti e la fornitura di meccanismi e formazione continua per prevenire il verificarsi di bias di automazione e preservare il giudizio umano autonomo[32].

2.2.1.6. Trasparenza e adeguata motivazione

61. Dovrebbe essere chiaramente documentato se l'IA generativa viene utilizzata in qualsiasi aspetto del processo decisionale giudiziario[33]. Le parti hanno il diritto di sapere quali elementi di una decisione giudiziaria sono stati assistiti da sistemi di IA generativa.

2.2.1.7. Diritto di accesso a un tribunale

62. Questo diritto fondamentale affronta potenziali compromessi attraverso sistemi automatizzati che sostituiscono o ostacolano l'accesso al processo decisionale giudiziario umano, influenzando il diritto al giudizio umano in questioni giudiziarie.
63. Le garanzie tecniche dovrebbero garantire un accesso significativo a un giudice persona fisica, confinare l'IA

generativa a funzioni strettamente assistive e vietare l'automazione totale delle decisioni giudiziarie[34].

2.2.2. Diritto alla libertà e alla sicurezza (Art. 5 CEDU)

64. Le decisioni che riguardano la libertà personale richiedono il massimo livello di supervisione e giustificazione umana. I sistemi di IA generativa che influenzano la detenzione, la custodia cautelare o le misure cautelari senza un'adeguata revisione umana e una giustificazione trasparente pongono rischi significativi per le libertà fondamentali[35].
65. L'introduzione dell'IA nelle funzioni giudiziarie dovrebbe seguire un approccio progressivo e incrementale, inizialmente limitato a compiti amministrativi o procedure standardizzate con risultati prevedibili. L'accuratezza, l'efficacia e la sicurezza dell'IA generativa dovrebbero essere rigorosamente valutate prima di espandersi a funzioni giudiziarie più sensibili.
66. Le garanzie tecniche dovrebbero includere sistemi di IA generativa che promuovono alternative alla detenzione, spiegazioni obbligatorie leggibili dall'uomo delle raccomandazioni generate dall'IA che riguardano la libertà e revisione umana completa di ogni decisione assistita dall'IA in cui potrebbe essere applicata la privazione della libertà.

2.2.3. Diritto alla privacy e alla protezione dei dati (Art. 8 CEDU)

67. La privacy e la protezione dei dati personali dovrebbero essere garantite se vengono utilizzati sistemi di IA generativa. Tutte le tecnologie e i dati impiegati dovrebbero rimanere sotto il controllo esclusivo della magistratura, al fine di prevenire accessi non autorizzati, fughe di dati o interferenze esterne e preservare la sovranità dell'istituzione sulle informazioni. I dati giudiziari dovrebbero rimanere sotto il controllo esclusivo delle autorità giudiziarie. I tribunali dovrebbero mantenere la capacità di cambiare fornitori di IA senza perdita di funzionalità nonché il diritto di verificare algoritmi e dati di addestramento attraverso processi di audit continui.
68. Le garanzie tecniche possono includere la pseudonimizzazione locale obbligatoria dei dati personali prima di caricare tali dati nel sistema di IA, l'esecuzione di modelli di IA su infrastrutture controllate dal governo o in infrastrutture commercialmente

disponibili; in quest'ultimo caso, mediate attraverso accordi contrattuali formali tra autorità pubbliche e fornitori di servizi tecnologici, e l'implementazione di rigorosi protocolli di protezione dei dati[36].

2.2.4. Diritto a un ricorso effettivo (Art. 13 CEDU)

- 69. I sistemi automatizzati alimentati dall'IA non dovrebbero ostacolare l'accesso effettivo ai rimedi o al controllo giurisdizionale[37].
- 70. Le garanzie tecniche dovrebbero garantire la revisione umana obbligatoria delle decisioni assistite dall'IA e stabilire procedure correttive specifiche per contestare determinazioni automatizzate o "influenzate dall'IA".

2.2.5. Divieto di discriminazione (Art. 14 CEDU e Protocollo n. 12)

2.2.5.1. Principio di non discriminazione

- 71. I bias algoritmici nei sistemi di IA generativa pongono rischi significativi di replicare o amplificare la discriminazione esistente attraverso caratteristiche protette, potenzialmente violando i principi fondamentali di uguaglianza[38].
- 72. Le garanzie tecniche dovrebbero includere audit periodici sui bias, pulizia e bilanciamento dei dati, termini di regolarizzazione dell'equità durante l'addestramento (come vincoli di parità demografica o quote equalizzate), test controfattuali/A-B e monitoraggio continuo delle metriche di discriminazione[39].

2.2.5.2. Neutralità e prevenzione attiva dei bias

- 73. I sistemi di IA dovrebbero essere periodicamente valutati per rilevare e mitigare i bias, garantendo che non perpetuino discriminazioni esistenti o creino nuove forme di disuguaglianza nel processo decisionale giudiziario. Il monitoraggio sistematico dei bias attraverso tutte le caratteristiche protette è obbligatorio.

2.2.6. Ulteriori garanzie

2.2.6.1. Spiegabilità e motivazione trasparente

74. I sistemi di IA utilizzati nella sfera giudiziaria dovrebbero fornire spiegazioni chiare e comprensibili del percorso seguito per ciascun risultato o raccomandazione, in modo che gli operatori giudiziari possano verificarli criticamente. A tal fine, i sistemi dovrebbero divulgare il loro processo di inferenza interna in modo trasparente e verificabile, garantendo un allineamento sostanziale ed evitando simulazioni di conformità superficiale che mascherano la loro reale logica[40]. I sistemi dovrebbero essere in grado di indicare le fonti specifiche di ciascuna affermazione generata. Dovrebbero essere impiegati solo modelli interpretabili che rivelano i loro processi decisionali interni.

2.2.6.2. Proporzionalità e cautela nell'implementazione dell'IA generativa

75. L'introduzione dell'IA nelle funzioni giudiziarie dovrebbe seguire un approccio progressivo e incrementale, inizialmente limitato a compiti amministrativi o procedure standardizzate con risultati prevedibili. L'accuratezza, l'efficacia e la sicurezza dell'IA generativa dovrebbero essere rigorosamente valutate prima di espandersi a funzioni giudiziarie più sensibili.

2.2.6.3. Responsabilità dello Stato per i danni causati dall'uso dell'IA generativa

76. Lo Stato dovrebbe assumersi la responsabilità per qualsiasi danno causato da errori giudiziari o guasti dei sistemi di IA generativa, in linea con i principi esistenti di responsabilità governativa. Né l'autonomia né la complessità tecnologica dell'IA generativa dovrebbero essere invocate per esentare lo Stato dal suo dovere di fornire risarcimento.

2.2.6.4. Principio di equità ed efficienza procedurale

77. L'introduzione dell'IA nell'amministrazione della giustizia non dovrebbe comportare un deterioramento delle garanzie procedurali o un pregiudizio della tutela giurisdizionale effettiva; al contrario, dovrebbe rafforzare i diritti procedurali e l'efficacia giudiziaria. Ad esempio, dovrebbe migliorare l'accesso a un tribunale, offrire un migliore accesso alle prove o migliorare i mezzi per presentare meglio il proprio caso. Inoltre, l'implementazione dell'IA generativa dovrebbe tradursi in

miglioramenti misurabili nei tempi di risposta, nella chiarezza della comunicazione, nella coerenza delle decisioni e nell'efficienza della gestione dei casi. I tribunali dovrebbero mantenere la capacità di tornare ai sistemi precedenti se l'IA generativa non migliora oggettivamente i servizi giudiziari. Valutazioni regolari dovrebbero valutare il rispetto di questo principio[41].

3. GLOSSARIO

Principio di neutralità tecnologica

È generalmente descritto come la libertà di individui e organizzazioni di scegliere la tecnologia più appropriata e adatta alle loro esigenze e requisiti per lo sviluppo, l'acquisizione, l'uso o la commercializzazione, senza dipendenze dalla conoscenza coinvolta come informazione o dati[42].

Regolarizzatori di equità durante l'addestramento del modello

Tecniche matematiche specifiche come:

- Regolarizzatori di parità demografica: penalizzano il modello quando i tassi di previsione differiscono significativamente tra i gruppi demografici
- Regolarizzatori di quote equalizzate: garantiscono tassi di veri positivi e falsi positivi simili tra i gruppi protetti
- Vincoli di equità individuale: garantiscono che individui simili ricevano un trattamento simile
- Debiasing adversarial: utilizzano reti adversarial per rimuovere informazioni sulle caratteristiche protette dalle rappresentazioni apprese

Note e riferimenti

[1] Javier Ercilla Garcia, Giudice, Spagna.

[2] Warankar, M., & Patil, R. (2024). Generative Artificial Intelligence. International Journal of Scientific Research in Engineering and Management, 8(04). doi:10.55041/IJSREM31146

- [3] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Chen, D., Dai, W., Chan, H. S., Madotto, A., & Fung, P. (2022). Survey of hallucination in natural language generation. arXiv. <https://arxiv.org/abs/2202.03629>
- [4] Wolters Kluwer. (2024). 2024 Future Ready Lawyer Survey Report: Legal innovation: Seizing the future or falling behind? Wolters Kluwer Legal & Regulatory. <https://www.wolterskluwer.com/en/know/future-ready-lawyer-2024>
- [5] n=33 intervistati su 43 Stati membri del Consiglio d'Europa interpellati. Decimali arrotondati a numeri interi.
- [6] Artificial Lawyer. (2025, 24 aprile). UK Courts roll out Microsoft Copilot for judges, update GenAI rules. <https://www.artificiallawyer.com/2025/04/24/uk-courts-roll-out-microsoft-copilot-for-judges-update-genai-rules/>
- [7] El País (2025, 22 settembre). Do you already know what carpeta.jusitcia.es is? Recuperato da <https://cincodias.elpais.com/extra/s/2025-09-22/ya-sabes-lo-que-es-carpetajusticiaes.html>
- [8] Weiser, B. (2023, 27 maggio). Here's what happens when your lawyer uses ChatGPT. The New York Times. <https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html>
- [9] MIT Sloan Teaching & Learning Technologies. (2024, 12 novembre). When AI gets it wrong: Addressing AI hallucinations and bias. <https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias/>
- [10] Carrick, D., & Kesteven, S. (2023, 24 giugno). This US lawyer used ChatGPT to research a legal brief with embarrassing results. We could all learn from his error. ABC News. <https://www.abc.net.au/news/2023-06-24/us-lawyer-uses-chatgpt-to-research-case-with-embarrassing-result/102490068>
- [11] Xu, Z., Jain, S., & Kankanhalli, M. (2024). Hallucination is Inevitable: An Innate Limitation of Large Language Models. arXiv. <https://arxiv.org/abs/2401.11817>

- [12] Nie, F., Yao, J.-G., Wang, J., Pan, R., & Lin, C.-Y. (2019, luglio). A simple recipe towards reducing hallucination in neural surface realisation. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 2673-2679). Association for Computational Linguistics.
<https://doi.org/10.18653/v1/P19-1256>
- [13] L'elenco seguente non è esaustivo.
- [14] Colelough, B. C., & Regli, W. (2025). Neuro-symbolic AI in 2024: A systematic review. arXiv. <https://arxiv.org/abs/2501.05435>
- [15] Felin, T., & Holweg, M. (2024). Theory Is All You Need: AI, Human Cognition, and Decision Making. Disponibile su:
<https://ssrn.com/abstract=4737265>.
- [16] Li, X., Cai, Z., Wang, S., Yu, K., & Chen, F. (2025). A survey on enhancing causal reasoning ability of large language models. arXiv.
<https://arxiv.org/abs/2503.09326>
- [17] Ibid.
- [18] Ercilla García, J. (2025). Justicia automatizada: entre las inteligencias artificiales que fingen y las que persuaden. Lex Social, Revista De Derechos Sociales, 15 (1), 1-39.
<https://doi.org/10.46661/lexsocial.11652>
- [19] Consiglio d'Europa. (2018). Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data [Convenzione 108 modernizzata], Consiglio d'Europa Trattati Serie - No. 223. Strasburgo.
- [20] Ad esempio, il Comitato Tecnico di Stato spagnolo per l'Amministrazione Elettronica della Giustizia (2024). Politica sull'Uso dell'Intelligenza Artificiale nell'Amministrazione della Giustizia. Una traduzione di cortesia in inglese può essere trovata su: <https://www.administraciondejusticia.gob.es/cteaje/normativa-complementaria>.
- [21] Ad esempio, il Regio Decreto spagnolo 311/2022, che regola lo Schema Nazionale di Sicurezza.

- [22] Commissione Europea per l'Efficienza della Giustizia (CEPEJ). (2023, 4 dicembre). Assessment Tool for the Operationalisation of the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment. Consiglio d'Europa.
- [23] Consiglio d'Europa, Comitato sull'Intelligenza Artificiale. (2024, 28 novembre). Methodology for the Risk and Impact Assessment of Artificial Intelligence Systems from the Point of View of Human Rights, Democracy and the Rule of Law (HUDERIA Methodology). Consiglio d'Europa.
- [24] Convenzione per la Protezione dei Diritti dell'Uomo e delle Libertà Fondamentali, 4 nov. 1950, ETS No. 5, 213 U.N.T.S. 221. https://www.echr.coe.int/documents/d/echr/convention_ENG
- [25] L'articolo 6 § 1 della CEDU garantisce il diritto a "(...) un'equa e pubblica udienza entro un termine ragionevole davanti a un tribunale indipendente e imparziale costituito per legge."
- [26] Livello di comprensione di come il sistema basato sull'IA sia arrivato a un determinato risultato, vedere ISO/IEC TR 29119-11:2020
- [27] Richiedere alle organizzazioni di mantenere registri dettagliati delle operazioni di IA, inclusi input di dati, decisioni del modello, errori e interazioni degli utenti, vedere ISO/IEC 42001:2023
- [28] Lo Strumento di Valutazione della CEPEJ identifica come rischio n. 8: "Rischio di vantaggio ingiusto per una parte nel processo".
- [29] Lo Strumento di Valutazione della CEPEJ identifica come rischio n. 11: "Rischio di generazione e uso di disposizioni legali inesistenti da parte dell'IA generativa".
- [30] L'articolo 6 § 1 CEDU richiede il giudizio da parte di un "tribunale indipendente e imparziale costituito per legge".
- [31] Salvi, F., Horta Ribeiro, M., Gallotti, R., & West, R. (2024). On the Conversational Persuasiveness of Large Language Models: A Randomized Controlled Trial. arXiv preprint arXiv:2403.14380
- [32] Lo Strumento di Valutazione della CEPEJ identifica come rischio n. 12: "Rischio di disempowerment e limitazione della responsabilità

del giudice attraverso l'uso di IA non spiegabile". Il framework HUDERIA consente la valutazione ricorrente dei potenziali impatti negativi sull'autonomia giudiziaria.

[33] Commissione Europea per l'Efficienza della Giustizia. (2025, 31 gennaio). Reflections of the AIAB on the use of artificial intelligence in judicial systems (CEPEJ-AIAB(2025)1Rev5).

[34] Lo Strumento di Valutazione della CEPEJ identifica come rischio n. 6: "Rischio che l'IA sostituisca l'accesso al giudice".

[35] L'articolo 5 § 1 CEDU protegge il diritto alla libertà soggetto solo a eccezioni rigorosamente definite e legali. L'articolo 5 § 2 stabilisce il diritto di essere informati "dei motivi del suo arresto". L'articolo 5 § 4 garantisce il diritto al controllo giurisdizionale sulla legalità della detenzione.

[36] Lo Strumento di Valutazione della CEPEJ identifica come rischio n. 2: "Rischio di divulgazione di dati personali o segreti commerciali". Il framework di analisi dell'impatto sulla privacy di HUDERIA affronta specificamente i rischi di divulgazione non autorizzata di dati sensibili attraverso la sua metodologia di analisi del rischio COBRA.

[37] L'articolo 13 CEDU stabilisce che "Ogni persona i cui diritti e libertà riconosciuti nella presente Convenzione siano stati violati, ha diritto a un ricorso effettivo davanti a un'istanza nazionale, anche quando la violazione sia stata commessa da persone agenti nell'esercizio delle loro funzioni ufficiali."

[38] L'articolo 14 CEDU stabilisce che "Il godimento dei diritti e delle libertà riconosciuti nella presente Convenzione deve essere assicurato senza distinzione di alcuna specie, come di sesso, di razza, di colore, di lingua, di religione, di opinione politica o di altro genere, di origine nazionale o sociale, di appartenenza a una minoranza nazionale, di ricchezza, di nascita o di altra condizione."

[39] Lo Strumento di Valutazione della CEPEJ identifica come rischio n. 10: "Rischio di discriminazione o amplificazione della discriminazione". Il processo di coinvolgimento degli stakeholder di

HUDERIA consente l'identificazione degli impatti differenziati sui gruppi vulnerabili.

[40] Greenblatt, R., Denison, C., Wright, B., Roger, F., MacDiarmid, M., Marks, S., Treutlein, J., Belonax, T., Chen, J., Duvenaud, D., Khan, A., Michael, J., Mindermann, S., Perez, E., Petrini, L., Uesato, J., Kaplan, J., Shlegeris, B., Bowman, S. R., & Hubinger, E. (2024). Alignment faking in large language models [arXiv preprint arXiv:2412.14093v2]. <https://doi.org/10.48550/arXiv.2412.14093>

[41] Gli indicatori di valutazione dovrebbero includere indagini di soddisfazione degli utenti, confronto delle metriche di prestazione con i sistemi pre-IA, feedback degli stakeholder da giudici, personale giudiziario, avvocati e cittadini, e valutazione dei miglioramenti delle tempistiche procedurali.

[42] Vedere: https://icannwiki.org/Technology_neutrality