



UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI DIRITTO PUBBLICO
ITALIANO E SOVRANAZIONALE



PoliS AI NEWS

Newsletter sull'Intelligenza Artificiale
a cura di PoliS-Lombardia

Anno II – n. 6/2025

In questo numero

In evidenza

Focus

Normativa

Applicazioni alla Pubblica Amministrazione

AI in pillole

Notizie

Commenti

Corsi, convegni e pubblicazioni

In questo numero

Che effetto ha l'AI sulla giustizia amministrativa? Ce lo mostra il Consiglio di Stato che, attraverso la partnership con Oracle, ha dato il via ad una piccola rivoluzione amministrativa che raccontiamo "In evidenza", insieme

ad un'analisi sull'efficacia di Julia (la chatbot del Comune di Roma); un report sui principali **abusi di ChatGPT**; una serie di **previsioni sui trend di sviluppo dell'AI**; e un'indagine su come **l'AI possa favorire l'inclusione delle persone con disabilità**. Nel "Focus" affrontiamo una questione solo a prima vista teorica: **gli LLM possono esprimere manifestazioni di pensiero?** In "Normativa" un resoconto sul **Ddl appena approvato in Senato**. La "pillola didattica" spiega **come garantire la trasparenza dell'AI nella Pubblica amministrazione**. Poi, come sempre, esempi, applicazioni alla PA, notizie, commenti... Buona lettura!

In evidenza

L'AI, il Consiglio di Stato e quella rivoluzione nella fase istruttoria

(a cura di A. Negrelli)

La trasformazione digitale che sta investendo **la Giustizia amministrativa italiana punta sull'Intelligenza Artificiale** per migliorare l'efficienza e la qualità delle attività. L'obiettivo: supportare i magistrati nello studio delle cause senza compromettere il loro ruolo decisionale. Grazie a questi sviluppi, **il Consiglio di Stato ha recentemente ricevuto il premio "Agenda Digitale 2024"** dall'Osservatorio della School of Management del Politecnico di Milano per il miglior progetto innovativo nella categoria delle Amministrazioni centrali, un riconoscimento che conferma il percorso di digitalizzazione avviato già nel 2017. Da febbraio 2024, [attraverso la partnership con Oracle](#), il Consiglio di Stato ha tracciato una strada che proietta il sistema giudiziario in una nuova era di efficienza, trasparenza e sostenibilità.

Il **linguaggio giuridico, intrinsecamente complesso e ambiguo, rappresenta una sfida** per la piena applicazione degli LLM, considerato oltretutto che i LLM funzionano elaborando contenuti basati su correlazioni probabilistiche tra parole, senza una comprensione reale delle relazioni causali.

I LLM, però, offrono potenzialità significative per la giustizia amministrativa e opportunità di trasformazione di processi complessi come la ricerca delle fonti normative, l'analisi dei fascicoli digitali e la redazione di provvedimenti giudiziari. Consentono **ricerche personalizzate, sintesi mirate e il confronto tra documenti giuridici**, migliorando efficienza e precisione. Nel supporto alla redazione di atti, possono **identificare criticità stilistiche e favorire chiarezza e sinteticità**, ampliando anche la comprensibilità delle decisioni per un pubblico non tecnico. Infine, nei metodi di risoluzione alternativa delle controversie, offrono strumenti per l'automazione delle analisi preliminari e la simulazione di scenari negoziali, segnando una possibile svolta nelle modalità di gestione stragiudiziale delle dispute.

Il Consiglio di Stato ha implementato **cinque applicazioni basate sull'AI**.

- 1. Richiamo e verifica delle fonti:** rilevazione automatica e immediata dei riferimenti giurisprudenziali presenti in un testo.
- 2. Identificazione di casi simili e ricorsi connessi:** utilizzo della ricerca semantica per individuare procedimenti simili ancora pendenti.
- 3. Collegamento tra atti e leggi di riferimento:** riconoscimento automatico delle normative pertinenti in base al contenuto di un atto.
- 4. Analisi delle sentenze:** individuazione di similitudini tra decisioni giurisprudenziali attraverso connessioni semantiche avanzate.
- 5. Anonimizzazione dei provvedimenti:** sviluppo di strumenti in grado di oscurare dati sensibili rispettando la normativa sulla *privacy* senza compromettere la comprensibilità del testo.

Il Regolamento UE 2024/1689 (AI Act) classifica **i sistemi di AI destinati alla Giustizia come "ad alto rischio"**, sottoponendoli a requisiti stringenti. L'applicazione dei LLM alla giustizia presenta, infatti, **tre rischi principali**:

l'**automation bias**, ovvero la tendenza a riporre una fiducia eccessiva nei suggerimenti forniti da un LLM, sottovalutando segnali divergenti e compromettendo così l'indipendenza decisionale umana; le **allucinazioni**, che consistono nel rischio di decisioni fuorviate da risposte non sempre supportate da dati reali; il **rischio di omologazione alla cultura giuridica del Common Law**, vale a dire l'appiattimento sulla cultura giuridica basata sul precedente giurisprudenziale come fonte normativa, che deriva dalla assoluta dominanza nella geopolitica dell'AI Generativa di Paesi come Usa, UK e Cina per l'addestramento e la progettazione dei LLM.

Un primo bilancio

All'[Oracle Cloud World Tour all'Allianz MiCo di Milano](#), **Brunella Bruno** (Consigliere di Stato, in servizio presso la VII Sezione Giurisdizionale e responsabile del Servizio per l'informatica del Consiglio di Stato) **ha raccontato così l'adozione dell'AI nella giustizia amministrativa italiana**: «Abbiamo realizzato un ecosistema interno, super controllato e vigilato, un'esclusiva della giustizia amministrativa perché realizzato con il contributo di tecnici e magistrati. (...) Questo pensando a diversi casi d'uso, ad esempio le intercettazioni, per agire in maniera rapida. Un sistema che preservi anche l'anonimizzazione, senza rendere difficile la lettura del testo. Pipeline e architetture dedicate per potersi adattare a cambiamenti rapidi, (...) **grazie ad una condivisione di saperi che rappresenta una leva positiva per il cambiamento**».

Una rivoluzione in corso, quindi. Che offre strumenti capaci di supportare il lavoro giuridico con una nuova "intelligenza collettiva". Ma che solleva domande sulla progettazione responsabile, il *prompt engineering* e la formazione. Anche perché all'orizzonte c'è il **rischio della Shadow AI, ossia l'uso non trasparente di LLM** non specificamente addestrati per il diritto amministrativo, con implicazioni su sicurezza, *privacy* e *bias* decisionali...

Per un approfondimento:

*[Il Consiglio di Stato investe sull'intelligenza artificiale per una giustizia più efficiente | Servicematica, 17 febbraio 2025](#)

*[Oracle Cloud World Tour 2025 e la rivoluzione digitale della Giustizia Amministrativa: tra intelligenza artificiale e sostenibilità | RivistaAI, 18 marzo 2025](#)

*[L'Intelligenza artificiale al servizio della Giustizia Amministrativa | Network 360, 14 febbraio 2025](#)

*[A.Caffo, L'AI tra infrastruttura e flessibilità, le novità dell'Oracle CloudWorld Tour 2025 | Datamanager Online, 18 marzo 2025](#)

*Il [sito](#) del Consiglio di Stato con le determine di affidamento annuale dei servizi tecnologici a Oracle

Julia, la chatbot del Comune di Roma, alla prova del Giubileo

Il **Comune di Roma ha lanciato** lo scorso 7 marzo **Julia**, una **chatbot progettata per aiutare ad esplorare e vivere la città** eterna, che funziona tramite chat (Facebook Messenger, WhatsApp e Telegram). Creata insieme a Microsoft e NTT Data, Julia **offre risposte personalizzate in modo intuitivo**, come Chat GPT, ai turisti in arrivo per il Giubileo (sono attese 35 milioni di visite).

Julia è **operativa 24/7, parla 80 lingue** e offre un aiuto rapido ai turisti disorientati. Dà informazioni su alberghi e ristoranti di qualità (grazie alle guide de *L'Espresso*) e sui servizi disponibili lungo i percorsi – come fontanelle, servizi igienici e infopoint turistici. Inoltre, fornisce indicazioni su farmacie aperte e di turno, Pronto soccorso e tempi di attesa, nonché su eventi culturali, spettacoli e mostre.

Ma come sta reggendo alla prova del Giubileo?

L'impressione è che diverse cose vadano messe a punto. [Uno dei problemi principali](#) riscontrati è la difficoltà di Julia nel comprendere il contesto di alcune richieste complesse. In parecchie occasioni **la chatbot ha fornito risposte non pertinenti o incomplete**: chiedendo, ad esempio, quale fosse la via più lunga di Roma, ha risposto la Via Appia (anziché via Cristoforo Colombo), oppure, se si richiede un percorso, tende a rimandare ad altre app, come Google Maps, invece di offrire risposte immediate.

Inoltre, Julia non offre informazioni sui bus notturni (problema serio per chi si muove per la prima volta in una città in cui due linee metro chiudono prima di mezzanotte ben quattro giorni alla settimana). Altri problemi li ha

dati con la toponomastica e **mancano, curiosamente, le segnalazioni di [eventi culturali](#) di punta** della città, come le mostre su Frida Kahlo e quella su Caravaggio.

Infine, Julia non è ancora in grado di distinguere correttamente tra richieste ordinarie e situazioni urgenti, causando ritardi nelle risposte in casi di emergenza. **C'è ancora da lavorare su vari aspetti**, ma intanto una nuova versione con servizi di bike e car sharing è attesa già a maggio.

*La [chat](#) di Julia

[A. Longo, Julia, il chatbot del Comune di Roma: come si usa, perché e i vantaggi. La nostra prova | Il Sole 24 Ore, 14 marzo 2025](#)

Che cosa succede se ChatGPT è usato dai regimi (e da chi sta dalla loro parte)

Un anno fa, OpenAI è stata la prima società a pubblicare **rapporti sugli abusi di ChatGPT** da parte di regimi autoritari per controllare i cittadini o minacciare altri Stati, o da parte di altri soggetti per operazioni di influenza occulta, truffe, spam e attività informatiche dannose. Il [rapporto](#), pubblicato a febbraio 2025, illustra alcuni casi studio sulle minacce neutralizzate di recente.

Le due principali attività bloccate coinvolgevano attori «che sembrano provenire dalla Cina». Con una prima operazione, denominata “*Peer Review*”, **una rete di account ChatGPT** è stata impiegata per **analizzare articoli in inglese sulle proteste nelle città occidentali per i diritti degli uiguri**, minoranza etnica musulmana che vive nel nord della Cina e subisce persecuzioni. Gli stessi profili di ChatGPT sono stati poi impiegati per **generare materiale promozionale per uno strumento di monitoraggio dei social media** utilizzato per identificare le discussioni politiche in Cina – soprattutto gli inviti a partecipare a manifestazioni sui diritti umani – da segnalare alle autorità. Gli operatori hanno usato i modelli di OpenAI anche all'estero (Stati Uniti, Germania e Regno Unito). **Gli account sono stati bloccati.**

Nell'altra operazione, chiamata *Sponsored Discontent*, **OpenAI ha bannato alcuni account ChatGPT impiegati per generare commenti critici in inglese su Cai Xia**, politologa cinese dissidente in esilio negli Usa. Gli stessi profili sono stati **usati per produrre articoli in spagnolo contro gli Usa, poi pubblicati a pagamento su vari media latinoamericani**, tra cui *La República* (Perù), *El Universal* (Messico), *Expreso* (Ecuador). Si tratta di una campagna di influenza più sofisticata, mai osservata in precedenza da OpenAI.

Un **altro abuso** interrotto sul nascere ha avuto **origine in Corea del Nord**: una serie di account di ChatGPT hanno creato CV, cover letters e profili LinkedIn falsi per **favorire assunzioni fraudolente** di lavoratori nordcoreani a servizio del regime. Durante i colloqui, i candidati usavano vari software per nascondere voce e posizione e ricorrevano a ChatGPT per generare risposte convincenti. Dopo aver ottenuto l'impiego, si avvalevano dell'AI per scambiare email con i colleghi – così da sembrare normali lavoratori e non destare sospetti– e per svolgere compiti tecnici, come la scrittura di codici; **gli stipendi venivano però trasferiti al regime**. Questa tattica era già stata segnalata da Microsoft e Google come parte di **un'operazione di finanziamento del regime di Pyongyang**: infiltrare il mercato del lavoro occidentale, **permettendo alla Corea del Nord di ottenere valuta estera bypassando le sanzioni internazionali.**

Altri casi segnalati nel report riguardano l'utilizzo di account ChatGPT per truffe romantiche, per generare strumenti di attacco informatico e malware e per tentativi di frode elettorale.

*il [testo](#) del report

I trend dell'AI, visti dalla Germania

Il panorama dell'AI è in continua evoluzione con impatti significativi sul lavoro, sull'economia, sull'istruzione e sulla governance. Il [report AI Trend 2025](#), realizzato da Statworx e AI Hub Frankfurt, individua **le principali traiettorie di sviluppo**, offrendo insight e raccomandazioni pratiche.

«**Gli agenti di AI rivoluzioneranno il mercato del lavoro**», è il primo trend evidenziato nel rapporto; non a caso anche [secondo i ricercatori di Gartner](#), «entro il 2028, circa il 15% delle decisioni lavorative quotidiane verranno prese autonomamente da agenti AI». A conferma di ciò, lo studio fa riferimento all'assistente AI di Klarna, il fornitore svedese di servizi di pagamento, che ha gestito due terzi delle chat coi clienti – ben 2,3 milioni di conversazioni – svolgendo il lavoro di 700 dipendenti e raggiungendo «identici livelli di customer satisfaction». «**Gli agenti di AI potrebbero automatizzare il 60-70% delle ore di lavoro nell'attuale economia globale**». Se da un lato questi agenti potranno contrastare la carenza di lavoratori qualificati, dall'altro lo studio evidenzia come entro il 2030 ci saranno 78 milioni di nuovi posti di lavoro, basati su nuove competenze.

L'altra grande tendenza evidenziata dal report riguarda proprio la formazione. L'AI Act ne ha introdotto l'obbligo: **le aziende e le PA dovranno garantire ai dipendenti una conoscenza adeguata dell'AI.** E ciò potrebbe accelerarne l'adozione nelle imprese europee, riducendo il divario con Cina e Stati Uniti. Viene sottolineato anche come **le piattaforme di apprendimento AI «democratizzeranno l'educazione»:** strumenti come ChatGPT, NotebookLM e Duolingo Max stanno personalizzando i percorsi di apprendimento. «L'apprendimento automatico potrebbe ridurre in modo massiccio i costi dell'istruzione in tutto il mondo e aumentare significativamente l'alfabetizzazione entro il 2030».

Il report sottolinea anche come «**l'intelligenza artificiale sia al centro di una rivoluzione che sta cambiando radicalmente campi come la medicina, la ricerca sui materiali e la scienza del clima**», proponendo vari esempi. Uno [studio statunitense](#) ha dimostrato un miglioramento del 21% delle diagnosi di cancro al seno grazie all'AI. Lo [strumento AI GNoME di Google](#) nel 2023 ha scoperto 380mila “materiali stabili” (quelli potenzialmente più utili per lo sviluppo di superconduttori e di fonti energetiche più efficienti – come batterie e celle solari). Alcuni [ricercatori cinesi](#) hanno sviluppato un catalizzatore alimentato dall'AI che è riuscito a produrre ossigeno usando «materiali provenienti da Marte», risparmiando «circa 2.000 anni di lavoro umano» e dimostrando l'efficienza dell'AI nel superare le sfide tecniche.

Infine, lo studio fa alcune **previsioni sul progresso tecnologico:** nel 2025, i “**Large Action Models**” (LAM) e i “**Computer User Assistants**” (CUA) – che combinano le capacità di elaborazione del linguaggio dei modelli LLM, con la capacità di prendere decisioni logiche e di eseguire autonomamente azioni – **gestiranno attività sul desktop dell'utente**, automatizzando processi complessi come l'organizzazione di file, la gestione delle mail e l'assistenza in compiti lavorativi (un esempio è [Rabbit R1](#), che oggi prenota appuntamenti e esegue azioni nel browser).

*Qui il [testo](#) del report

Come rendere l'AI un fattore di inclusione per le persone con disabilità

L'intelligenza artificiale facilita l'integrazione delle persone con disabilità sul posto di lavoro o rischia di diventare un ostacolo alla loro inclusione? Del tema si occupa [un report](#) pubblicato dal **Center for Democracy & Technology in collaborazione con l'American Association of People with Disabilities**, che propone una serie di raccomandazioni rivolte a policymakers e Pubblica amministrazione. L'obiettivo? **Assicurarsi che l'innovazione tecnologica proceda di pari passo con la tutela dei diritti fondamentali, con particolare attenzione a quelli delle persone con disabilità.**

Questa sfida è particolarmente complessa nel mondo del lavoro. I [dati dello U.S. Bureau of Labor Statistics](#) mostrano che, sebbene **nel 2023 il tasso di occupazione delle persone con disabilità negli Usa abbia raggiunto livelli record** (il 36,1% per le donne e 38,2% per gli uomini), **il divario rispetto ai loro omologhi senza disabilità**

resta significativo: nello stesso anno, il tasso di occupazione della popolazione senza disabilità è stato del 70,3% per le donne e del 79,7% per gli uomini.

I rischi più seri riguardano i **bias** nella valutazione. Già oggi, almeno [un terzo delle aziende statunitensi utilizza sistemi di AI per la selezione del personale](#): il pericolo è che algoritmi di valutazione dei curriculum non calibrati escludano candidati con percorsi lavorativi non lineari. Altrettanto **rischioso è l'uso di tecnologie di riconoscimento facciale e vocale**: già controverse di per sé, nei casi di disabilità potrebbero interpretare erroneamente i movimenti oculari delle persone cieche o la cadenza vocale di una persona neurodivergente, o fraintendere i movimenti facciali di un lavoratore con la Sindrome di Tourette.

Un altro fronte critico è quello del "bossware" – software di monitoraggio costante dell'attività lavorativa –: il ricorso a tecnologie invasive può generare stress e discriminazioni, penalizzando in modo sproporzionato chi necessita di pause per gestire la propria condizione.

Un **modello efficace** per mitigare questi rischi potrebbe ispirarsi al [Maryland's Artificial Intelligence Governance Act](#) del 2024. Sebbene questa legge non menzioni esplicitamente i diritti dei disabili, introduce l'identificazione dei sistemi di AI "ad alto rischio" – quelli che potrebbero influenzare negativamente diritti civili, pari opportunità, accesso a risorse critiche o privacy –, imponendo una **valutazione d'impatto prima della loro implementazione**.

L'appello lanciato dagli esperti è chiaro: audit accurati, scambio di risorse tecniche e buone pratiche condivise vanno affiancate al **coinvolgimento delle comunità per i diritti delle persone con disabilità**. Solo così l'AI può trasformarsi in una leva d'inclusione.

* Il [testo](#) del report

Focus



LLM - Diritti - Libertà di espressione

Ma gli LLM possono esprimere manifestazioni di pensiero? E vanno moderati?

In [questa clip video](#) con il prof. Marco Bassini dell'Università di Tilburg ci interroghiamo se i contenuti prodotti dai sistemi di intelligenza artificiale generativa possano beneficiare di una tutela costituzionale ed essere equiparati a manifestazioni di pensiero. La domanda, apparentemente teorica, nasconde implicazioni pratiche di grande importanza, specialmente in considerazione del confine molto labile tra interventi tecnici volti a prevenire la generazione di output inaccurati, falsi o addirittura diffamatori e possibili derive censorie.

Normativa

Senato – DDL sull'AI

Il 20 marzo il Senato ha approvato un **disegno di legge (atto n. 1146, Senato) su “Disposizioni e deleghe al Governo in materia di intelligenza artificiale”**, importante passo in avanti verso l'adozione di una normativa nazionale che integri le previsioni del regolamento europeo sull'AI (l'AI Act). In armonia con la regolamentazione dell'Unione europea, **il testo propone una disciplina generale sull'impiego, lo sviluppo e la sperimentazione di sistemi di AI**, con l'obiettivo di promuoverne un utilizzo sicuro e responsabile.

Il disegno di legge sottolinea fin dai primi articoli il carattere antropocentrico dell'AI, che non deve mai sostituirsi all'uomo, ma piuttosto essergli di supporto. Viene evidenziata la necessità di garantire trasparenza, sicurezza, affidabilità e inclusività nello sviluppo e nell'uso dei sistemi intelligenti, richiedendo al tempo stesso che questi siano accessibili anche alle persone con disabilità. Una **particolare attenzione è riservata alla cybersicurezza**, considerata imprescindibile lungo tutto il ciclo di vita dei sistemi di AI per tutelare i cittadini da utilizzi impropri o manipolazioni dei sistemi. **Altrettanta rilevanza rivestono i principi in materia di informazione, tutela dei dati personali e sviluppo economico.**

Il disegno di legge individua anche alcune **disposizioni settoriali**. **L'ambito sanitario** è uno dei più sensibili e, per questo, la proposta prevede che **l'AI possa essere utilizzata per migliorare le attività di prevenzione, diagnosi e cura** – a patto che non vengano mai compromessi i diritti dei pazienti. I sistemi di AI non potranno decidere al posto dei medici, ma solo offrire previsioni e suggerimenti, nel rispetto della deontologia e della libertà professionale. Parallelamente, si promuove l'uso dell'AI per facilitare la vita quotidiana delle persone con disabilità, favorendo la loro autonomia e inclusione sociale.

Anche il **mondo del lavoro** è profondamente interessato dalla trasformazione tecnologica. **L'intelligenza artificiale** viene vista come uno strumento in grado di **migliorare le condizioni lavorative, aumentare la produttività e garantire la sicurezza dei lavoratori**. Tuttavia, l'uso dell'AI non potrà mai violare la dignità della persona o introdurre forme di discriminazione. A tale proposito, viene istituito un Osservatorio presso il Ministero del Lavoro con il compito di monitorare l'impatto dell'AI sull'occupazione e promuovere la formazione.

La **Pubblica amministrazione** e il settore giudiziario vengono a loro volta incoraggiati ad **adottare sistemi di AI per migliorare l'efficienza** dei servizi e ridurre i tempi delle procedure, ma sempre sotto la guida di personale umano responsabile delle decisioni.

Per rafforzare la competitività nazionale, la legge prevede che **lo Stato investa fino a un miliardo di euro nel sostegno a piccole e medie imprese innovative nei settori dell'AI, della cybersicurezza e delle tecnologie emergenti come il calcolo quantistico**. Le politiche pubbliche puntano inoltre a valorizzare la ricerca e la collaborazione tra università, centri tecnologici e imprese. Il disegno di legge dedica una particolare attenzione alla **formazione**: si promuove l'alfabetizzazione digitale in ambito scolastico e universitario, con l'obiettivo di formare nuove competenze.

Sul fronte della **tutela dei diritti** viene rafforzata la protezione dei dati personali, soprattutto per i minori, e si stabilisce che solo le opere che siano frutto del contributo creativo umano possano godere della protezione del diritto d'autore. Vengono inoltre introdotte nuove sanzioni penali per reati commessi con l'ausilio dell'intelligenza artificiale, come la diffusione di *deepfake*.

La **governance dell'intero sistema è affidata a due autorità nazionali**: l'Agenzia per l'Italia digitale, responsabile della promozione e regolazione dell'AI, e l'Agenzia per la cybersicurezza nazionale, incaricata del controllo e della vigilanza.

* Il [testo](#) del disegno di legge

(a cura di Marco Bassini, Assistant Professor di Fundamental Rights and Artificial Intelligence, Tilburg University)

Applicazioni alla Pubblica Amministrazione

ITALIA

TALIA SF- Puglia

L'app, sviluppata da Innova-Puglia (società in-house della Regione Puglia), raccoglie e organizza una grande quantità di documenti provenienti dai progetti finanziati dal Programma Interreg MED.

Serve a gestire le politiche di innovazione e valorizzare i risultati dei progetti più riusciti.

[TALIA SF - Extract and organize knowledge coming from the analysis of a huge quantity of textual documents | AI-X](#)

Predictive Justice – Sant'Anna di Pisa

Progetto che mira a sviluppare una metodologia di analisi del materiale giurisprudenziale per «gestire flussi continui di grandi basi di dati»

www.predictivejustice.eu

MONDO

UPCAST Project – Gestione dati

Il progetto UPGCAST mira a sviluppare plug-in per facilitare lo scambio di dati tra aziende, pubbliche amministrazioni e cittadini. Questi strumenti si

UNIONE EUROPEA

Una serie di casi studio, suddivisi per ambito di applicazione, nei vari Paesi dell'Ue

[AI and GenAI adoption by local and regional administrations | European Committee of the Regions](#)

Polpo – Paesi Bassi e Belgio

Piattaforma completamente automatizzata che monitora Parlamento, Governo e Consiglio di Stato.

Gli utenti possono creare ricerche e ricevere aggiornamenti in tempo reale su proposte legislative, mozioni, documenti e interrogazioni parlamentari.

[Polpo- Your political monitor Powered by Artificial Intelligence](#)

concentrano su settori chiave come la gestione e lo scambio dei dati, la privacy, la monetizzazione e la negoziazione automatizzata, garantendo il rispetto delle normative europee sull'intelligenza artificiale e delle procedure etiche.

[UPCAST Project](#)

AI in pillole

Le garanzie di trasparenza dell'AI nelle Pubbliche Amministrazioni

(a cura di *Annalisa Negrelli*)

Un numero crescente di settori pubblici ad alto rischio (finanza, sanità, forze dell'ordine) si affida a modelli AI per il processo decisionale. **Migliorare la conoscenza dei cittadini su come vengono addestrati questi modelli e su come determinano i risultati generati** crea fiducia nelle decisioni di AI e nelle amministrazioni che le utilizzano. Per **trasparenza dell'AI** si intende proprio quella procedura di esplicazione del suo funzionamento che **aiuta gli utenti a comprendere come sia stato creato un sistema di intelligenza artificiale e come questi modelli prendano le decisioni**. Questa procedura va oltre la semplice spiegazione dei processi decisionali dell'AI, comprendendo altresì i fattori legati allo sviluppo dei sistemi AI e alla loro implementazione, documentando e condividendo ad esempio la logica e il ragionamento dell'algoritmo di AI sottostante, gli input di dati utilizzati per addestrare il modello, i metodi impiegati per la valutazione e la convalida del modello e altro ancora.

Ma come garantire questa trasparenza?

Sebbene la trasparenza dell'AI differisca in base al caso d'uso, all'organizzazione e al settore (pubblico o privato), **esistono alcune strategie che devono essere tenute in considerazione quando si creano sistemi di AI, e che** includono l'adozione di principi chiari di fiducia e trasparenza e la loro integrazione nell'intero ciclo di vita dell'AI, oltre alla loro **divulgazione completa in ogni fase del ciclo di vita dell'AI**.

A questo proposito, sono disponibili piattaforme software e strumenti che possono aiutare ad automatizzare la raccolta di informazioni e altre attività di [governance dell'AI](#).

Per garantire tale divulgazione, le amministrazioni pubbliche devono stabilire quali **informazioni** condividere tra quelle elencate:

- Nome del modello
- Scopo
- Livello di rischio
- Modello di politica
- Generazione del modello
- Dominio previsto
- Dati di addestramento
- Accuratezza nei test e nella formazione
- Distorsione
- Metriche di resistenza agli attacchi
- Metriche di equità
- Metriche di cosiddetta "spiegabilità"
- Recapiti

Per quanto riguarda poi **le modalità di presentazione dei dati**, le PA possono presentare informazioni sulla trasparenza dell'AI **in vari formati, come documenti stampati o video**. Il formato dipende sia dal destinatario, sia dal caso d'uso, e possono consistere in un documento in evoluzione, modellato sulla dichiarazione di conformità del fornitore (*SDoC*); una *pagina ufficiale di politica* sul sito Web della PA che descriva nel dettaglio come sta mettendo in atto iniziative di AI trasparenti; *risorse didattiche* come documenti e video per aiutare gli utenti a capire come viene utilizzata l'AI; strumenti di sollecitazione di un *dibattito pubblico* dal punto di vista etico attraverso attività ufficiali di pubbliche relazioni, eventi, social media e altre comunicazioni; *documenti di ricerca e altre comunicazioni* basate su dati AI per offrire insight sull'uso, lo sviluppo e i vantaggi dell'AI in quel determinato settore pubblico.

La disciplina normativa

La disciplina che regola l'uso dell'AI è in continua evoluzione. L'**Eu AI Act** indica specifici **obblighi di trasparenza per tipi di AI**: in particolare, i sistemi di AI destinati a interagire con le persone dovrebbero essere progettati per informare gli utenti che stanno interagendo con un sistema di AI, a meno che ciò non sia ovvio per l'individuo dal contesto; e ancora, i sistemi AI che generano testo o immagini dovrebbero contrassegnare gli output come generati o manipolati dall'AI.

Anche a livello globale sono presenti discipline sulla trasparenza AI, come i **Blueprint for an AI Bill of Rights**, un insieme di cinque principi e pratiche associate per aiutare a guidare la progettazione, l'uso e l'implementazione dei sistemi AI – il quarto, denominato "Avviso e spiegazione", riguarda direttamente la trasparenza. Il **Quadro politico globale del Processo di Hiroshima sull'AI**, lanciato nel 2023 in seguito allo sviluppo del Vertice del G7 di Hiroshima, invita le organizzazioni alla «pubblicazione di rapporti sulla trasparenza» e la «condivisione responsabile delle informazioni».

La *trasparenza* dell'AI (come è stato creato il modello, quali dati lo hanno addestrato e come prende le decisioni?), poi, è naturalmente correlata ad altri due concetti: la *spiegabilità* (come è arrivato il modello a questo risultato?) e l'*interpretabilità* (come prende le decisioni?)

I problemi

Le pratiche di AI trasparenti, però, sollevano anche problemi di sicurezza e privacy. Più informazioni vengono fornite sul funzionamento interno di un progetto di AI, **più facile sarà per gli hacker individuare e sfruttare le sue vulnerabilità**. Un'altra sfida alla trasparenza è rappresentata dal **compromesso tra trasparenza e protezione della proprietà intellettuale**: fino a che punto è possibile disvelare i processi di addestramento e decisione della AI senza danneggiare la proprietà intellettuale del software? E ancora, come superare la difficoltà di spiegare in modo chiaro programmi complessi e intricati e algoritmi di apprendimento automatico (come [le reti neurali](#)) a persone non esperte? Come ovviare alla mancanza di standard di trasparenza a livello globale per l'AI?

[A. Jonker, A. McGrath, A. Gomstyn, *Che cos'è la trasparenza AI?* | IBM, 6 settembre 2024](#)

[S. Mancini e E. Maran, *Trasparente, spiegabile e interpretabile: solo così l'IA farà bene a tutti* | Agenda Digitale, 21 marzo 2024](#)

[F. Paglieri, *Quale trasparenza per l'IA?* | Il Mulino, 23 febbraio 2024](#)

Notizie

[D. Casati, *Google, arrivano in Italia le AI Overview: ecco come cambia il motore di ricerca. E si pensa già alla «fase 2»* | Corriere della Sera, 26 marzo 2025](#)

[R. Cosentino, L'intelligenza artificiale arriva su WhatsApp: come funziona e cosa si può fare | Corriere della Sera, 25 marzo 2025](#)

[M. Schiaffino, L'intelligenza artificiale può cambiare il mercato delle imprese in Europa | Wired, 24 marzo 2025](#)

[Minori adescati in rete con l'IA, allarme dell'associazione Meter | La Repubblica, 22 marzo 2025](#)

[M. Garofalo, Deregulation dell'Intelligenza artificiale secondo Trump | Corriere della Sera, 19 marzo 2025](#)

[P. Feliciani, OpenAI e Anthropic sfidano le regole Ue: ecco gli impatti | Agenda Digitale, 19 marzo 2025](#)

[G. Esperti, Intelligenza artificiale, c'è un altro comitato in Italia. Quello dell'Agcom | Wired, 19 marzo 2025](#)

[P. Albanese, Anna Luce D'Amico candidata sindaca a Taranto, è la prima creata con l'intelligenza artificiale | La Repubblica, 19 marzo 2025](#)

[L. Mischitelli, Distillazione dell'intelligenza artificiale: perché sarà centrale nello sviluppo | Agenda Digitale, 19 marzo 2025](#)

[Tumore al seno per 1 donna su 8: l'Intelligenza artificiale è «amica» | Il Sole 24 Ore, 19 marzo 2025](#)

[Più della metà delle aziende italiane destinerà il 5% del fatturato all'Intelligenza artificiale | La Repubblica, 17 marzo 2025](#)

[Il Foglio lancia il Foglio AI, primo quotidiano dell'intelligenza artificiale | Il Sole 24 Ore, 17 marzo 2025](#)

[G. Rusconi, Intelligenza artificiale in azienda: ceo entusiasti ma impreparati a gestire il cambiamento | Il Sole 24 Ore, 17 marzo 2025](#)

[C. Metz, K. Wiese, M. Hernandez, M. Isaac, A. Singhvi, How A.I. Is Changing the Way the World Builds Computers | New York Times, 16 marzo 2025](#)

[V. M. Affer, D. Delfino, Le linee guida del Vaticano sull'IA: etica e innovazione | Agenda Digitale, 13 marzo 2025](#)

Commenti

[L. Tremolada, L'intelligenza artificiale può aiutare la ricerca scientifica? Forse anche troppo | Il Sole 24 Ore, 24 marzo 2025](#)

[F. Fubini, I social rendono stupidi? L'intelligenza artificiale, l'innovazione e il ritardo dei giovani: il caso-Italia | Corriere della Sera, 24 marzo 2025](#)

[C. Walters-Davies, For great stories, we need people, not AI | The Guardian, 19 marzo 2025](#)

[R. Staglianò, Chi perde e chi guadagna con l'IA | La Repubblica, 19 marzo 2025](#)

[M. Carmignani, *Intelligenza artificiale generale, cos'è davvero? La verità oltre l'hype* | Agenda Digitale, 18 marzo 2025](#)

[C. Maurizio, *Intelligenza artificiale nell'istruzione, così prepara al mondo del lavoro* | Agenda Digitale, 18 marzo 2025](#)

[F. Santelli, *Portare l'IA nell'industria. Così si contrasta il dominio americano* | La Repubblica, 17 marzo 2024](#)

[G. Iuvinale, N. Iuvinale, *IA: il piano cinese per dominare il mercato globale entro il 2030* | Agenda Digitale, 17 marzo 2025](#)

[L'Europa accelera sull'AI con le nuove AI Factories, ma l'Italia può contare su CINECA ed ENI | Rivista AI, 17 marzo 2025](#)

[AI e il paradosso della ricerca: strumenti avanzati, risultati errati e sfide per l'informazione | Rivista AI, 17 marzo 2025](#)

[J. Naughton, *Now you don't even need code to be a programmer. But you do still need expertise* | The Guardian, 16 marzo 2025](#)

[K. Roose, *Powerful A.I. Is Coming. We're Not Ready* | New York Times, 14 marzo 2025](#)

[M. McArdle, *The case for how AI can make us feel more human - not less* | The Washington Post, 11 marzo 2025](#)

Corsi, convegni e pubblicazioni

Corsi

[Artificial intelligence for business- Master part time | Italian Tech Academy](#)

[AI Survival Kit | Il Sole 24 Ore](#)

Eventi e convegni

[RomeCup 2025 - What's next? Intelligenza umana e artificiale. Le sfide per il benessere olistico | Università degli Studi di Roma Tre e Campidoglio, 7-9 maggio 2025](#)

[Convegno dei risultati di Ricerca dell'Osservatorio Smart City | Polimi School Of Management, 7 maggio 2025](#)

Pubblicazioni

[G. Lo Sapió, *L'intelligenza artificiale generativa nella giustizia amministrativa* | 3 febbraio 2025](#)

[A. Amidei e U. Ruffolo, *Diritto dell'Intelligenza Artificiale \(Volume I\)*, Luiss University Press](#)

[A. Aresu, *Geopolitica dell'intelligenza artificiale*, Feltrinelli](#)

[L'intelligenza artificiale nel permitting 5.0 | IAIA 2025 Bologna Italy, 2 maggio 2025](#)

Strumenti

@MetaAI

Dai prossimi giorni basterà cercare l'icona con un cerchio blu su Facebook, Messenger, Instagram e WhatsApp e avviare una chat intelligente disponibile

in sei lingue europee, tra cui **l'italiano**. Tra le funzionalità più interessanti- oltre a quella di avere **un assistente a disposizione** in tutte e quattro le app- c'è la possibilità di usare Meta AI all'interno dei **gruppi WhatsApp**, digitando **@MetaAI seguito da una domanda**.

[G. Taviani, *Meta AI arriva in Italia: come funziona l'intelligenza artificiale integrata su Instagram, WhatsApp e Facebook* | Corriere della Sera, 20 marzo 2025](#)

Link attivi al 28 marzo 2025

Prodotto da: PoliS-Lombardia

Coordinamento editoriale a cura di Davide Perillo

Comitato Scientifico: **Marco Sica, Marco Bassini, Annalisa Negrelli**

(hanno collaborato: Beatrice Capitanio e Annaclara De Tuglie)