



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. 15/2025

In questo numero

In evidenza

Focus

Normativa

Applicazioni alla Pubblica Amministrazione

AI in pillole

Notizie

Commenti

Corsi, convegni e pubblicazioni

In questo numero

La scorsa settimana è stata approvata la **legge italiana sull'AI**: ne approfondiamo contenuti e portata "In Evidenza". Poi, partendo da un'indagine della Bocconi, facciamo il punto della situazione **sull'adozione dell'AI**

nelle strutture sanitarie lombarde: più della metà (il 57%) non ha ancora fatto il primo passo. Altri temi: l'uso dell'AI negli appalti pubblici, una ricerca dell'Università di Stanford sul suo impatto nel mondo del lavoro (a risentirne di più sono i giovani...) e Diella, l'avatar digitale inserito dal governo albanese nella propria squadra ministeriale per vigilare sulla gestione dei fondi pubblici. Il "Focus" è dedicato al boom dei data center (anche in Italia). In "Normativa", uno sguardo alla vicenda giudiziaria che ha visto coinvolto il colosso Anthropic per la violazione del diritto d'autore. E con la "Pillola didattica" entriamo nel merito dell'Explainable XAI: cos'è, perché è importante, e in che settori è necessaria? Poi, come sempre, esempi, applicazioni alla PA, notizie, commenti... Buona lettura!

In evidenza

Anche l'Italia ha la sua legge sull'AI (di Marco Bassini)

La scorsa settimana, il Parlamento ha approvato una legge quadro sull'Intelligenza artificiale per regolare in modo chiaro e responsabile l'uso di queste tecnologie in tutti i settori della società. La legge – fortemente intrecciata con il Regolamento europeo sull'AI – segna un passaggio fondamentale: da una visione "tecnica" dell'AI a una visione "politica" e "sociale", che mette al centro diritti, trasparenza, equità, sicurezza e sviluppo. La legge disciplina una serie di aspetti che il regolamento ha rimesso alla discrezionalità dei legislatori degli Stati membri, con particolare riguardo alla governance e alla definizione delle autorità competenti. Una breve rassegna sintetica.

Principi generali

La legge promuove un uso dell'AI antropocentrico, cioè incentrato sull'essere umano; responsabile e trasparente, specialmente nell'utilizzo di dati e algoritmi; sicuro e inclusivo, anche per le persone con disabilità; coerente con la Costituzione e il diritto Ue, soprattutto per tutelare i diritti e le libertà fondamentali.

Settori prioritari di intervento

L'AI viene disciplinata con un'attenzione particolare per alcuni settori strategici:

- **sanità**: l'AI può supportare diagnosi e cure, ma senza sostituire il ruolo del medico come agente umano. Viene favorita la ricerca (anche su dati sensibili) per finalità di salute pubblica, con misure rilevanti di protezione dei dati, in modo da non penalizzare lo sviluppo di sistemi AI da utilizzare in ambito medico;
- **lavoro**: viene promosso un uso che aumenti qualità e sicurezza del lavoro, evitando discriminazioni; è istituito un Osservatorio nazionale sul lavoro e l'AI;
- **Pubblica amministrazione**: l'AI potrà semplificare processi e migliorare i servizi pubblici, ma senza sostituire la decisione umana;
- **giustizia**: è previsto l'uso per semplificare l'organizzazione interna, ma le decisioni restano esclusivamente in capo ai magistrati;
- **professioni intellettuali**: l'AI può supportare il lavoro di avvocati, medici, architetti, ma non sostituirlo. I clienti devono essere sempre informati in modo chiaro.

Dati, algoritmi e cybersicurezza

Vengono introdotte misure per disciplinare l'uso di dati e algoritmi per l'addestramento dei modelli AI; per potenziare la cybersicurezza nazionale attraverso l'AI e, infine, per creare spazi di sperimentazione controllata (le cosiddette regulatory sandboxes) per lo sviluppo di nuove tecnologie.

Strategia nazionale e governance

Nasce una "Strategia nazionale per l'AI", coordinata dalla Presidenza del Consiglio, che punta a:

- incentivare **ricerca, innovazione e trasferimento tecnologico**;
- promuovere **investimenti (fino a 1 miliardo di euro) in start-up** e imprese innovative nei settori AI, cybersicurezza e calcolo quantistico;
- rafforzare la **collaborazione tra il settore pubblico e quello privato**.

Aspetti penali e tutela del diritto d'autore

Vengono introdotte **nuove fattispecie penali per contrastare l'uso illecito dell'AI** come:

- la **diffusione di contenuti deepfake** creati con l'ausilio dell'AI
- l'uso dell'AI per **commettere reati**
- l'estrazione non autorizzata di dati da opere protette dal diritto d'autore (in coordinamento con le previsioni dell'AI Act e della Direttiva Copyright europea).

Inoltre, la legge **individua le autorità competenti nell'Agenzia per l'Italia Digitale (AGID) e nell'Autorità per la cybersicurezza nazionale (ACN)**, rispettivamente designate come autorità di notifica e come autorità di vigilanza del mercato. Il governo poi, attraverso vari decreti legislativi, dovrà definire i poteri di queste autorità, anche di carattere sanzionatorio. Sono individuate in **IVASS, Banca d'Italia e CONSOB le autorità di vigilanza del mercato** per quanto attiene ai sistemi impiegati in ambito finanziario e assicurativo. E **rimangono validi i ruoli** delle altre autorità con rilevanti poteri in materia, come il Garante privacy e l'Autorità per le garanzie nelle comunicazioni nella sua veste di coordinatore per i servizi digitali ai sensi del Digital Services Act.

Molti **aspetti operativi non sono ancora stati definiti** e sono rimandati a decreti legislativi che il Governo dovrà emanare entro 12 mesi.

*Il [link](#) al testo della legge

*[Un commento video](#)

L'AI nella sanità lombarda: innovazioni in crescita, ma l'adozione resta limitata

A che punto siamo con l'adozione dell'AI in ambito medico in Lombardia? [Un'indagine della Bocconi](#) fotografa **l'uso delle applicazioni di Intelligenza artificiale all'interno di 46 strutture pubbliche e private della regione**. Più precisamente, sono state mappate 21 ASST, 4 IRCCS e 6 ATS per il pubblico, e 15 enti tra ospedali, IRCCS e ambulatori per il privato. Di questi, solo 20 usano l'AI in ambito clinico: **più della metà (il 57%) non ha ancora fatto il primo passo**.

Tra le **applicazioni censite, 56 in totale**, il 54% supporta la diagnosi, il 48% la prognosi e solo l'11% riguarda l'ottimizzazione dei trattamenti. **È la radiologia l'area con il maggior numero di casi d'uso (30%)**, seguita da oncologia e diabetologia.

«La maggior parte delle applicazioni è ancora legata alla diagnostica per immagini, dove l'offerta commerciale è ampia e consolidata», [spiega Giulia Cappellaro](#), professoressa associata di Public Management alla Bocconi. In effetti, lo studio evidenzia come **il 75% delle applicazioni AI venga usato proprio per elaborare immagini e segni**, oltre che per analizzare dati sanitari strutturati – come cartelle cliniche elettroniche o percorsi clinico-terapeutici personalizzati. «Le applicazioni più innovative, ad esempio per la prognosi in pazienti cronici, sono ancora in fase di sviluppo».

Queste ultime vengono realizzate soprattutto dagli **IRCCS, privati e a vocazione di ricerca, che stanno investendo per elaborare soluzioni proprie**, anche grazie a partnership accademiche o tecnologiche. Le applicazioni diagnostiche, invece, sono più diffuse all'interno delle strutture che acquistano apparecchiature commerciali CE-marked (cioè, conformi ai requisiti Ue) già disponibili sul mercato: sono 14 le organizzazioni che scelgono questa strategia, con una prevalenza del pubblico (67%).

Quelli che sviluppano sistemi AI al proprio interno (in totale 6 organizzazioni) sono anche **gli unici ad istituire strutture organizzative formali per governare l'adozione dell'AI**. Gli altri, invece, si affidano a iniziative informali o ad hoc.

Lo studio indaga anche i **principali ostacoli all'adozione dell'AI**. Secondo gli sviluppatori interni, sono la **privacy, l'interoperabilità e la mancanza di incentivi economici**; mentre le strutture che comprano soluzioni commerciali citano la **cultura organizzativa e la scarsa fiducia degli operatori**. I "non-adopters", invece, ritengono che la barriera principale sia **la carenza di personale qualificato**. Ma a investire nella formazione sono in pochi: solo il 33% delle strutture del primo gruppo offre corsi specifici.

Fotografia finale: una **realtà sfaccettata** in cui quello che serve «è un investimento nella collaborazione tra strutture sanitarie, università, fornitori tecnologici e istituzioni: senza un ecosistema coeso, l'innovazione rischia di restare confinata a pochi centri».

Per approfondire

*il [link](#) allo studio

[B.Orlando, Sanità lombarda e intelligenza artificiale: una rivoluzione a metà corsa | Bocconi Hub News ed Eventi, 25 luglio 2025](#)

Così l'AI può aiutare gli appalti pubblici

Il 5 settembre [un Protocollo d'intesa siglato presso il Ministero delle Infrastrutture e Trasporti](#) ha reso finalmente operativa la previsione **dell'articolo 30 del nuovo Codice appalti sull'uso dell'Intelligenza artificiale** per le stazioni appaltanti. Cosa prevede il testo? Al centro dell'accordo vi è la collaborazione tra Ministero, giustizia amministrativa ed enti di settore per **condividere banche dati e competenze, con l'obiettivo di sperimentare tecnologie di AI nelle procedure di gara**. Dal punto di vista operativo le novità sono disponibili per le Stazioni appaltanti sulla sezione [DigitAPP](#) della piattaforma/Hub Contratti Pubblici.

Grazie al protocollo d'intesa, la **giurisprudenza amministrativa in materia di contratti pubblici – come sentenze e pareri – verrà raccolta, elaborata e integrata tramite algoritmi AI** all'interno di DigitAPP. Questo permetterà di sviluppare anche altri strumenti avanzati di consultazione normativa e supporto decisionale per le stazioni appaltanti, tra cui **motori di ricerca giuridica basati su AI, assistenti virtuali per i RUP e dataset arricchiti** utili alla pubblicazione di linee guida operative.

Tra le sperimentazioni più rilevanti spicca il progetto [AI4RUP](#), un assistente digitale intelligente in fase di test sulla piattaforma DigitApp. Attivo dal luglio 2025, AI4RUP consente ai Responsabili Unici del Procedimento di **consultare in modo semantico il Codice dei contratti, ricevere risposte immediate e puntuali su questioni procedurali e gestire con maggiore efficienza le attività di gara** e contrattuali. Il chatbot, integrato nel portale MIT, rappresenta uno dei primi esempi di AI generativa applicata alla consulenza giuridica nella Pubblica amministrazione, semplificando il lavoro dei funzionari e garantendo aggiornamento costante e affidabilità nelle risposte.

Queste novità fanno seguito anche ad una [recente sentenza del Consiglio di Stato \(n. 4857/2025\)](#) che ha sancito la **piena legittimità dell'uso di sistemi di Intelligenza artificiale come strumenti di mero supporto nelle procedure di gara, a condizione che restino centrali il controllo e la decisione finale umani**. Questa sentenza rispecchia la necessità, per chi lavora nell'amministrazione pubblica, di un uso consapevole dell'AI negli appalti, nel rispetto di trasparenza, sicurezza e discrezionalità tecnica.

AI e lavoro: a rischiare sono i giovani (lo dicono i canarini...)

L'Intelligenza artificiale generativa sostituirà davvero gli esseri umani nel mercato del lavoro? Al dibattito estenuante che si è sviluppato in materia, basato finora soprattutto su stime e previsioni, si aggiunge (finalmente) una ricerca **fondata su dati reali**. E conferma le visioni più pessimistiche. **La sostituzione è già in atto e colpisce soprattutto i giovani fra i 22 e i 25 anni**. Lo studio [Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence](#), pubblicato ad agosto dall'Università di Stanford e riferito al mercato statunitense, ha analizzato i cedolini di milioni di lavoratori.

Il riferimento ai **canarini, usati nelle miniere di carbone come allarme perché più sensibili degli umani ai gas tossici, indica che si tratta di primi segnali**. Ma sono trend verificati su larga scala: **il campione conta fra i 3,5 e i 5 milioni di lavoratori ogni mese e offre la possibilità di osservare le dinamiche quasi in tempo reale**, basandosi sulle elaborazioni di ADP, la più importante società fornitrice di software per buste paga.

Gli accademici di Stanford si concentrano sulle tendenze del mercato del lavoro, distinguendo vari casi in base all'esposizione all'AI e per fasce d'età. L'analisi evidenzia **consistenti cali dell'occupazione per i lavoratori all'inizio della carriera: fino al 13%, nei settori più esposti all'AI**, come lo sviluppo software e il supporto clienti. Un dato in contrapposizione rispetto alla crescita dell'occupazione complessiva negli Usa, stimata fra il 6 e il 9%. Il fenomeno colpisce in particolare le figure di cui è possibile automatizzare il lavoro, come i responsabili marketing e vendite.

Perché i giovani sono la categoria più colpita? Gli studiosi ipotizzano che, possedendo una conoscenza quasi esclusivamente scolastica, possano essere più facilmente sostituiti. **I colleghi più anziani, al contrario, grazie alla conoscenza maturata con l'esperienza, sembrano essere più tutelati**.

Ma è davvero colpa dell'Intelligenza artificiale? Il calo di occupazione giovanile, in effetti, coincide con l'evoluzione dell'AI: [entrambi si verificano a partire dalla fine del 2022](#). **Per i giovani fra la fine del 2022 e il luglio 2025 il calo è quasi del 20%**, mentre nello stesso periodo il trend per le altre fasce d'età è in continua crescita. Una tendenza analoga emerge considerando i dati per settore. Nelle professioni informatiche e tra gli impiegati dei servizi, più esposti all'AI, si verifica un calo del 6% per chi ha fra i 22 e i 25 anni. Nei settori meno esposti all'AI, al contrario, si registra una crescita occupazionale generalizzata fra il 6 e il 13%.

Per evitare che questa tendenza possa portare effetti permanenti sull'occupazione giovanile, lo studio raccomanda **formazione continua per i giovani, incentivi alle imprese che li assumono e, soprattutto, di proseguire nella raccolta di dati aggiornati sull'impatto dell'AI**, per fasce di età e per i differenti settori, così da poter adottare interventi efficaci e tempestivi.

*Il [link](#) allo studio

E intanto, in Albania, l'AI va al governo...

Per la prima volta in Europa – e forse nel mondo – un **governo ha inserito un'Intelligenza artificiale nella propria squadra ministeriale**. È successo in **Albania**, dove il premier **Edi Rama**, al quarto mandato, ha presentato **Diella, un avatar digitale** chiamato a vigilare sulla **gestione dei fondi pubblici e sugli appalti**, con l'**obiettivo dichiarato di ridurre a zero i rischi di corruzione**. Una sfida particolarmente significativa in un Paese che, [secondo il Corruption Perceptions Index 2024 di Transparency International](#), si colloca all'**80° posto su 180 nazioni**.

Il nome scelto significa "sole" e la sua rappresentazione grafica richiama gli abiti tradizionali albanesi. Al di là del simbolismo, però, il compito di Diella è operativo: **analizzare enormi quantità di dati, rilevare anomalie e segnalare rapidamente eventuali irregolarità nella spesa statale**. Tra le sue funzioni ci sarà anche **l'ottimizzazione dei servizi ai cittadini**, grazie all'automazione delle pratiche amministrative.

Alla base di questa scelta c'è l'idea che un **sistema digitale, privo di interessi personali e legami politici, possa garantire maggiore trasparenza e imparzialità** rispetto agli strumenti tradizionali di controllo. È proprio su questo aspetto che Rama ha voluto insistere durante la conferenza nazionale del Partito socialista, occasione in cui ha presentato la nuova squadra di governo.

La decisione ha acceso il **dibattito interno e internazionale**. Alcuni osservatori la interpretano come un passo che potrebbe trasformare l'Albania in **laboratorio di innovazione istituzionale**, mentre altri sollevano **dubbi sulla reale capacità di un algoritmo di cogliere la complessità del contesto sociale e di proteggere appieno i diritti dei cittadini**.

Anche **l'opposizione ha reagito con durezza, contestando la legittimità dell'operazione** e sottolineando che un'intelligenza artificiale [non dispone né di cittadinanza né di personalità giuridica](#).

* Per approfondire

[S. Epifani, La ministra albanese creata con l'intelligenza artificiale rischia di trasformarsi in strumento di propaganda a scapito dei cittadini | Wired, 15 settembre 2025](#)

[C. Da Rold, In Albania nasce il primo ministro al mondo creato dall'intelligenza artificiale | Il Sole 24 Ore, 12 settembre 2025](#)

Focus



Ma quanto vale (e quanto inquina) la corsa ai data center?

La corsa ai dati sta trasformando i **data center in veri e propri asset strategici** per la competitività digitale ed economica. Ma non è solo questione di tecnologia: **la vera sfida riguarda la sostenibilità energetica**.

Secondo il Position Paper [L'Italia dei data center. Energia, efficienza, sostenibilità per la transizione digitale](#), realizzato da **TEHA Group con A2A** e presentato al Forum di Cernobbio, la **Data Economy italiana vale oggi 60,6 miliardi di euro** (2,8% del Pil), collocando il Paese al quarto posto in Europa (+UK), dopo Regno Unito, Germania e Francia. In uno scenario di pieno sviluppo, **entro il 2030 il settore potrebbe arrivare a 207 miliardi**, generando fino a **150 mila nuovi posti di lavoro entro il 2035**.

Il cuore di questa rivoluzione batte in Lombardia, con **Milano che si afferma come hub emergente in Europa**: dei **168 data center italiani**, con una potenza complessiva di 513 MW, **oltre il 60% si trova nella regione** e solo l'**area metropolitana milanese concentra 238 MW** di potenza, **pari al 46% del totale nazionale**, più di Madrid e il doppio di Zurigo.

Ma l'espansione dei data center ha un prezzo: i consumi. **Oggi queste infrastrutture assorbono quasi il 2% della domanda elettrica nazionale, che potrebbe salire fino al 13% entro il 2035**. [Secondo una recente analisi di Orlando \(Istituto per la Competitività\) e Sileo \(Fondazione Eni Enrico Mattei\)](#), la **maggior parte dei data center italiani è oggi collegata alla media o bassa tensione**, trattandosi ancora di impianti di dimensioni contenute. Ma anche l'Italia, in linea con il trend globale, **si sta muovendo verso strutture molto più grandi, capaci di assorbire tra 100 e 1.000 MW**, equivalenti ai consumi di centinaia di migliaia di famiglie.

Per gestire **la sfida della decarbonizzazione**, il paper di TEHA indica **quattro leve chiave**: recupero del calore, contratti di lungo termine per rinnovabili (PPA), riciclo dei RAEE e utilizzo di aree industriali dismesse (aree brownfield). Solo queste ultime, che in Italia coprono 3,7 milioni di m², equivalgono a un risparmio di suolo pari a 84 campi da calcio. **Integrando queste misure, si potrebbero evitare quasi 6 milioni di tonnellate di CO₂ all'anno e generare benefici economici stimati in 1,7 miliardi di euro**.

Gli esempi non mancano in Lombardia: a **Brescia il data center Qarnot** sarà integrato con la rete di teleriscaldamento urbano, garantendo calore a **1.350 appartamenti**; a **Milano, dal 2026, il progetto Avalon 3** recupererà oltre 15 GWh di energia termica l'anno, sufficienti a servire **1.250 famiglie in più** e ad **evitare l'emissione di 3.300 tonnellate di CO₂**. Secondo le stime di TEHA, **nella sola area metropolitana il recupero del calore dai data center vicini alle reti di teleriscaldamento – esistenti o in fase di sviluppo – sarebbe sufficiente a riscaldare oltre 530 mila famiglie**, con una **riduzione delle emissioni paragonabile all'installazione di 2,3 milioni di pompe di calore**.

La Regione, intanto, si conferma **laboratorio di buone pratiche anche sul fronte normativo**: in arrivo una **proposta di legge che introduce criteri ambientali e urbanistici per la localizzazione dei data center**. Anche la **Città Metropolitana di Milano** ha già predisposto strumenti innovativi, come la **"Tavola dell'Inversione Pianificatoria"** e la **"Mappa degli Elementi Utili"**, per guidare le scelte dei Comuni tenendo conto di fattori cruciali come il valore paesaggistico, il rischio idrogeologico e la presenza di reti ecologiche.

* Il link al [rapporto](#)

Normativa

Caso Anthropic: l'accordo da record che segna il futuro del copyright nell'AI

Anthropic ha accettato di pagare **1 miliardo e mezzo di dollari per transare la causa** avviata da oltre 500 mila autori **sull'uso non autorizzato di opere pirata per l'addestramento del chatbot AI Claude**. Si tratta del più grande accordo in un caso di copyright negli Stati Uniti e potrebbe influenzare il futuro dei modelli linguistici. La

transazione, che deve essere ancora ratificata dal giudice, si applica solo a quanto fatto da Anthropic nel passato e quindi non validerà eventuali usi futuri di opere letterarie da parte della società per il training dell'AI.

Sebbene questa vicenda legale non si chiuda con una sentenza, certifica comunque **alcuni segnali importanti**.

La causa verteva sull'**uso di libri piratati** (i cui contenuti, cioè, erano stati acquisiti in modo illegale) **per formare i modelli di AI di Anthropic**. Per questi particolari contenuti, proprio perché piratati, non esiste alcuna possibilità di "fair use" (quella dottrina anglosassone che forma una sorta di eccezione all'esclusiva del diritto d'autore). Il fair use **può essere applicato solo al materiale ottenuto in modo lecito**. In quei casi, nell'utilizzo di quelle opere letterarie per addestrare modelli di AI vi è un uso, per così dire, "trasformativo", e quindi non necessariamente illegale. Ma, appunto, qui non è possibile, a causa della provenienza.

L'Europa guarda con grande interesse a quanto accade negli Stati Uniti, non solo perché i principali soggetti di mercato sono statunitensi, ma anche **perché queste cause segnano un indirizzo significativo**. Le future controversie in materia di AI e diritto d'autore saranno sicuramente influenzate da questo caso e le società ne dovranno tenere conto nell'addestramento dei chatbot AI, optando per soluzioni più rigorose e trasparenti.

Per approfondire

[D. Di Geronimo, La causa tra Anthropic e gli autori USA potrebbe cambiare per sempre l'industria AI | Tutto Tech, 23 settembre 2025](#)

Applicazioni alla Pubblica Amministrazione

ITALIA

XLAW

In Italia il software XLAW analizza i trend criminali e i luoghi maggiormente soggetti a reati, con un'accuratezza compresa tra l'87 e il 93% nel predire l'avvenimento di nuovi crimini.

[XLAW® - Presentazione](#)

[X-Law: Polizia Predittiva – Rivista AI](#)

UNIONE EUROPEA

Chatbot NOA - Francia

Nella regione francese dell'Île-de-France è stato introdotto un assistente digitale che guida le imprese e le start-up attraverso le procedure burocratiche, riducendo tempi e complessità amministrative.

[Feedback forum de NOA](#)

DiVa Portal - Norvegia

In Norvegia, un sistema di AI è in grado di prevedere quali imprese presentano il rischio più elevato di insolvenza o di fallimento, consentendo l'intervento preventivo delle autorità locali.

[FULLTEXT01.pdf](#)

MONDO

Grok for Government – USA

Il modello Grok 4 è una suite di strumenti AI sviluppata da xAI, per portare un modello linguistico avanzato all'interno delle agenzie federali, dei ministeri e delle amministrazioni locali americane.

[Grok for Government | xAI](#)

AI in pillole

Explainable XAI: ovvero, come ti spiego l'AI

L'Intelligenza artificiale spiegabile (**Explainable XAI**) è un insieme di processi innovativi che consentono agli utenti di comprendere, e quindi considerare affidabili, i risultati generati dagli algoritmi di machine learning, deep learning e dalle reti neurali.

Questi possono essere raggruppati in 3 categorie:

- **best practice**, basate su tecniche usate da tempo dai data scientist per aiutare gli utenti a capire come i modelli usano i dati per i loro processi di apprendimento;
- **principi di design**, finalizzati a progettare dei sistemi AI più semplici da comprendere e implementare;
- **strumenti più efficienti e semplici da comprendere** che generano modelli di apprendimento sempre più raffinati, anche grazie alla possibilità di condivisione e sviluppo da parte della community.

Tutte queste tecniche vengono usate per descrivere un modello AI, ma anche gli effetti predeterminati e i potenziali *bias* e determinano l'accuratezza, l'equità, la trasparenza e i risultati del modello.

In altri termini, a differenza dell'interpretabilità della AI (che è la misura in cui un osservatore può comprendere il motivo di una decisione e dunque la percentuale di successo di un output AI), la **spiegabilità compie un ulteriore passo avanti per analizzare come l'AI sia giunta a quel determinato risultato.**

Perché è essenziale spiegare l'AI?

Con l'avanzare della tecnologia, diventa sempre più difficile comprendere il modo in cui l'algoritmo ottiene un determinato risultato. L'intero **processo di calcolo si trasforma spesso in una cosiddetta "black box"** che, a volte, nemmeno gli ingegneri o i data scientist riescono a comprendere. **Le reti neurali impiegate nel deep learning sono tra le più difficili** da capire. I **bias** – spesso legati a razza, sesso, età o provenienza geografica – sono **da sempre un rischio** quando si addestra un modello di AI. Inoltre, le prestazioni del modello possono essere soggette a **deviazioni perché i dati di produzione differiscono da quelli di addestramento.**

Per tutti questi motivi è fondamentale monitorare e gestire di continuo i modelli di addestramento al fine di promuovere la spiegabilità dell'AI, misurando al contempo l'impatto dell'utilizzo di tali algoritmi.

Vantaggi

In termini di **efficienza**, gli sviluppatori possono verificare che il sistema funzioni come previsto non solo per **garantire la conformità agli standard normativi**, ma anche per poter contestare o modificare i risultati, **riducendo così i rischi di non conformità, legali, di sicurezza e reputazionali** legati all'AI in produzione.

L'AI spiegabile è quindi fondamentale per creare un **ambiente di fiducia e sicurezza.**

Ma come si spiega l'AI?

La configurazione delle tecniche di XAI si articola in tre metodi principali. Da un lato, l'**accuratezza** e la **tracciabilità** delle previsioni soddisfano i requisiti tecnologici, mentre dall'altro la **comprensione** delle decisioni soddisfa le esigenze umane.

1. **Accuratezza della previsione:** la tecnica più usata è la cosiddetta **LIME** (*Local Interpretable Model-Agnostic Explanations*, spiegazione locale agnostica al modello), utilizzata per semplificare modelli complessi approssimandoli su piccole porzioni di dati con modelli interpretabili come la regressione lineare. Fornisce spiegazioni per singole predizioni, evidenziando le feature che influenzano maggiormente il risultato.
2. **Tracciabilità:** segna le limitazioni al modo in cui si possono prendere decisioni e la definizione di un ambito più ristretto per quanto riguarda regole e funzioni del machine learning. Un esempio di tecnica di tracciabilità della XAI è **DeepLIFT** (*Deep Learning Important Features*), che confronta l'attivazione di ogni neurone con il suo neurone di riferimento e mostra un collegamento tracciabile tra ogni neurone attivato e persino le dipendenze tra di essi.
3. **Comprensione delle decisioni:** molti utenti non si fidano dell'AI, ma per utilizzarla in modo efficiente devono imparare a farlo. Questo ambiente di fiducia si ottiene istruendo il "team umano" che lavora con l'AI, in modo che possa capire come e perché l'AI prende le decisioni.

Una tecnica piuttosto diffusa di XAI è poi quella del **proxy modeling**, che consiste nel creare una differente versione del modello, la cui struttura è dichiaratamente più semplice rispetto alla black box vera e propria. Si tratta di modelli in grado di approssimare e descrivere in maniera semplificata la notevole complessità elaborativa tipica del machine learning, ma al tempo stesso del tutto differenti rispetto a quelli in uso.

Casi d'uso dell'Explainable AI

- **Guida autonoma:** la XAI è fondamentale per spiegare ai passeggeri il funzionamento di questi sistemi, così da sentirsi al sicuro a bordo dei mezzi di trasporto.
- **Riconoscimento facciale:** le tecnologie basate sul riconoscimento dell'immagine, se applicate alle persone fisiche (come nei sistemi di videosorveglianza), comportano implicazioni relative a privacy e discriminazione, che rendono auspicabile un approccio il più possibile trasparente.
- **Sanità:** questi metodi di XAI consentono di accelerare la diagnostica, l'analisi delle immagini, l'organizzazione delle risorse, oltre che di migliorare la trasparenza e la tracciabilità del processo decisionale per la cura dei pazienti, o ancora di semplificare l'iter di approvazione dei prodotti farmaceutici.
- **Fintech:** la XAI applicata ai servizi finanziari migliora l'esperienza del cliente con un processo di approvazione di prestiti e crediti trasparente. Velocizza le valutazioni di rischio di credito, di gestione patrimoniale e di reato finanziario. Accelera la risoluzione di potenziali reclami e problemi.
- **Giustizia penale:** questi metodi consentono di ottimizzare i processi di previsione e valutazione del rischio, ma anche di accelerare le risoluzioni grazie all'impiego dell'AI spiegabile per l'analisi del DNA, della popolazione carceraria e per la previsione dei reati.
- **Telecomunicazioni:** spiegare come funzionano le reti, i protocolli e come vengono utilizzati i dati è fondamentale per rispettare le normative vigenti e generare fiducia nei cittadini.

Per approfondire:

[F. La Trofa, Explainable AI: cos'è, quali sono i principi e gli esempi | Tech4Future, 27 dicembre 2021](#)

[Cos'è l'AI spiegabile \(XAI\)? | IBM](#)

[A. Plebe, Explainable AI: perché pretendere spiegazioni dall'Intelligenza Artificiale è giusto ma non basta | Agenda Digitale, 3 marzo 2022](#)

Notizie

[A. Longo, Anche LinkedIn vuole addestrare i modelli di intelligenza artificiale con i nostri dati | Il Sole 24 Ore, 24 settembre 2025](#)

[L. Angelini, Prezzi bassi ma 900 milioni di utenti: perché l'India è la nuova terra di conquista per i big dell'AI | Corriere della Sera, 23 settembre 2025](#)

[M. Gaggi, La sfida dell'italiana Domyn con l'intelligenza artificiale «sovrana» per rivoluzionare la finanza | Corriere della Sera, 22 settembre 2025](#)

[E. Dusi, Un virus creato con l'IA: grazie a lui combatteremo i batteri resistenti agli antibiotici | La Repubblica, 21 settembre 2025](#)

[I. Frish, What Wall Street Sees in the Data Center Boom | The New York Times, 20 settembre 2025](#)

[E. Spagnuolo, I lavori che l'intelligenza artificiale non riuscirà a sostituire nei prossimi 10 anni | Corriere della Sera, 19 settembre 2025](#)

[C. Ponti, L'Italia ha la sua legge sull'AI, gli impatti privacy e cyber: tutti i punti chiave | Cyber Security 360, 18 settembre 2025](#)

[E. Capilupi, Come l'intelligenza artificiale entra nelle scuole secondo le nuove linee guida del ministero dell'Istruzione | Wired, 18 settembre 2025](#)

[K. Robison, ChatGPT e adolescenti, le nuove funzioni di OpenAI per proteggerli | Wired, 17 settembre 2025](#)

[S. Carmignani, Così l'AI prevede il rischio di oltre mille malattie. Con 10 anni di anticipo | La Repubblica, 17 settembre 2025](#)

[Copilot entra nella Camera dei Rappresentanti degli Stati Uniti | Rivista AI, 17 settembre 2025](#)

[V. Bansal, Sottopagati, precari e licenziati senza preavviso: le terribili condizioni delle persone che lavorano all'AI di Google | Wired, 16 settembre 2025](#)

Commenti

[P. Benanti, Il modello britannico: formare l'umano per godere dei vantaggi dell'AI | Il Sole 24 Ore, 24 settembre 2025](#)

[M. Suleyman, Necessaria una norma contro la ricerca di una coscienza illusoria | Il Sole 24 Ore, 24 settembre 2025](#)

[L'intelligenza artificiale entra nel gioco della previsione geopolitica | Rivista AI, 22 settembre 2025](#)

[M. Sideri, *Qualche domanda su salute e Intelligenza artificiale* | Corriere della Sera, 20 settembre 2025](#)

[Microsoft Fairwater Wisconsin: la corsa al Data Center AI più potente del mondo | Rivista AI, 20 settembre 2025](#)

[J. Ross, *L'intelligenza artificiale ha bisogno di molta più energia \(e non solo\)* | Start Magazine, 20 settembre 2025](#)

[P. Haski, *La Cina farà a meno dei processori statunitensi per l'ia* | Internazionale, 18 settembre 2025](#)

[H. Long, *Welcome to the era of stagflation-lite* | The Washington Post, 17 settembre 2025](#)

[A. Puliafito, *Le intelligenze artificiali a scuola* | Internazionale, 17 settembre 2025](#)

[F. Ognibene, *L'apprendistato, i poveri, l'Intelligenza artificiale: l'agenda del Papa* | Avvenire, 16 settembre 2025](#)

[R. Luna, *Intelligenza artificiale, avanti ma con giudizio* | Corriere della Sera, 16 settembre 2025](#)

[D. Guastella, *AI will make the rich unfathomably richer. Is this really what we want?* | The Guardian, 16 settembre 2025](#)

Corsi, convegni e pubblicazioni

Corsi

[Avantgrade, *Workshop: GEO AI Search. Primi nelle risposte AI* | Milano, 2 ottobre 2025](#)

[24 Ore Business School, *Corso Artificial Intelligence for Leaders* | Roma e Online, 3 ottobre 2025](#)

[24 Ore Business School, *Master Big Data, Intelligenza Artificiale e Business Analytics* | 24 ottobre 2025](#)

[Scuola FQ di Cittadinanza, *Intelligenza artificiale e potere* | 29 settembre- 11 novembre 2024](#)

Eventi e convegni

[The Innovation Group, *Milano Digital Week* | Milano, 1- 5 ottobre](#)

[Federchimica, *Giornata della ricerca chimica – L'Intelligenza Artificiale al servizio della ricerca chimica* | Milano, 2 ottobre 2025](#)

[ADM, SOGEI e Politecnico di Milano, *Ital-IA* | Milano, 8-9 ottobre 2025](#)

Pubblicazioni

[M. R. Taddeo, *Codice di guerra. Etica dell'intelligenza artificiale nella difesa* | Raffaello Cortina Editore, 2025](#)

[M. Iklé, A. Kolonin, M. Bennett, *Artificial General Intelligence* | Springer Nature, 2025](#)

[C. B. Frey, *How Progress Ends: Technology, Innovation, and the Fate of Nations* | Princeton University Press, 2025](#)

[G. Galli – A. Valfré, *L'Intelligenza Artificiale al servizio della Pubblica Amministrazione* | Università Cattolica del Sacro Cuore, 2025](#)

Link attivi al 26 settembre 2025

Prodotto da: PoliS-Lombardia

Coordinamento editoriale a cura di **Davide Perillo**

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

(hanno collaborato: Beatrice Capitanio, Annaclara De Tuglie, Chiara Rizzo, Vanna Toninelli)