



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 III – n. 7/2026

In questo numero

In evidenza

Focus

Normativa

Applicazioni alla Pubblica Amministrazione

AI in pillole

Notizie

Commenti

Corsi, convegni e pubblicazioni

## In questo numero

L'Intelligenza artificiale può fare davvero la differenza nel contrasto all'evasione fiscale? Ne parliamo nella sezione "In Evidenza" raccontando una serie di *best practices* di vari Paesi, insieme ad un approfondimento

sul **ruolo dell'AI in ambito farmaceutico** e all'analisi delle **trasformazioni che stanno investendo la ricerca** negli Stati Uniti. Chiude la sezione una **panoramica sulla figura dei data worker, lavoratori invisibili e sottopagati**, senza i quali le piattaforme AI non potrebbero funzionare. Con il "Focus" entriamo nel merito di **due sentenze molto discusse** che segnano un possibile **cambio di paradigma nella responsabilità delle piattaforme digitali, spostando l'attenzione sul design** invece che sui contenuti. In "Normativa", una lettura dell'**impatto sulle imprese italiane del Digital Omnibus e della Legge italiana sull'AI**. La "Pillola" è dedicata all'**Emergent Behaviour AI**, e cioè tutti quelle **abilità inaspettate**, tra cui il pericoloso fenomeno del **sycophancy**: cosa succede quando gli agenti AI compiaciono gli utenti? Poi, come sempre, esempi, applicazioni alla PA, notizie e commenti. Buona lettura!

## In evidenza

### AI Tax Administration: tra efficienza amministrativa e garanzia dei dati sensibili

di Annalisa Negrelli

Come emerge da un ampio studio dell'Ocse di qualche mese fa dal titolo [Governing with Artificial Intelligence](#), l'Intelligenza artificiale sta assumendo un ruolo sempre più rilevante **anche per supportare le PA nella lotta all'evasione fiscale**. In Europa, le amministrazioni tributarie sono tra le prime strutture pubbliche ad aver adottato sistemi di AI in grado di raccogliere, confrontare e analizzare grandi quantità di dati: **il 79% ha implementato sistemi di AI in ambito fiscale contributivo**, con lo scopo di contrastare le frodi, modernizzare i processi interni, migliorare l'erogazione dei servizi ai contribuenti e rafforzarne la capacità decisionale; **il 55% ha adottato assistenti virtuali e chatbot**; **il 45% ha introdotto sistemi di supporto** agli uffici fiscali nell'adozione delle **decisioni amministrative**.

Le tecniche più comuni oggi impiegate in questi ambiti spaziano **dall'analisi predittiva, ai modelli di machine learning, fino all'uso di immagini satellitari** per individuare, ad esempio, immobili non dichiarati al catasto. L'Ocse segnala inoltre un **uso crescente dell'AI per contrastare frodi massive, come richieste di rimborsi falsi generate automaticamente da reti criminali**.

Più in particolare, **in Francia**, l'amministrazione fiscale utilizza l'algoritmo AI della piattaforma [Géoportail](#), basato su un sistema di matching automatico tra immagini aeree e dichiarazioni catastali, alla ricerca di immobili non dichiarati. La **Grecia** ha sviluppato [IAPR | AAAE](#), un modello AI per individuare costruzioni non registrate, dotato inoltre di funzioni predittive per l'evasione IVA e le frodi sui carburanti. In **Austria** il [Predictive Analytics Competence Center](#), articolato in più divisioni (delle quali la Tax Analytics Division è dedicata proprio al contrasto dell'evasione fiscale), nel giro di pochi mesi dalla sua istituzione aveva già consentito un recupero di più di 180 milioni di euro e di **150 milioni solo nel 2025**. Strumenti di lotta all'evasione sono diffusi anche in [Portogallo](#) (chatbot fiscali, strumenti di pre-compilazione automatica con modelli predittivi, analisi comportamentale per prevenire errori) e [Spagna](#) (machine learning per individuare frodi IVA, analisi semantica di documenti, sistemi di clustering per controlli mirati).

**In Italia**, dopo una fase iniziale più cauta, l'Agenzia delle Entrate sembra intenzionata a farne largo uso nel 2026, come spiega nel suo [podcast Raffaello Lupi](#), professore di Diritto tributario a Roma "Tor Vergata". L'utilizzo di sistemi di AI nelle indagini patrimoniali e fiscali, come [Cerebro](#), è diretto a potenziare l'**accertamento analitico** ad esempio nelle **frodi carosello IVA**, sofisticati sistemi di evasione IVA su scala internazionale, oppure nei **falsi crediti d'imposta**. In questi ambiti l'algoritmo, a differenza dei meccanismi di controllo tradizionali e degli studi di settore, **non mira tanto alla stima dei ricavi sommersi dei contribuenti, ma a scovare anomalie documentali e flussi illeciti complessi, tramite l'analisi di banche dati integrate**. Sul

fronte della compliance tributaria, infatti, l'AI è strutturalmente **imbattibile** nell'intercettare schemi fraudolenti complessi, anche attraverso l'analisi di dati non strutturati come testi, documenti scansionati o contenuti pubblici sui social media.

Oltre a questi modelli con funzione repressiva, **i sistemi AI in materia fiscale svolgono un ruolo crescente nell'assistenza ai contribuenti**. Chatbot evoluti, come [IRAS Bot](#) dell'autorità fiscale di **Singapore**, rispondono in tempo reale a quesiti complessi, accedono a servizi autenticati e riducono notevolmente i tempi di attesa. In **Australia**, [ATO](#) "Alex" e altri sistemi di pre-compilazione e rilevazione di anomalie avvisano il contribuente in live chat, prevenendo errori e contenziosi. In questo senso, anche [CRA](#), live chatbot **canadese**, KAI, l'assistente virtuale **neozelandese**, e il modello **statunitense** IRS supportano i contribuenti in ambito fiscale e tributario.

L'adozione di sistemi AI nel fisco solleva, tuttavia, **alcune questioni giuridiche delicate**. Il caso olandese dell'[AI Toeslagenaffaire \(Dutch Childcare Allowance Scandal\)](#), che nel 2021 aveva costretto alle dimissioni l'allora primo ministro Mark Rutte, rappresenta un monito sui **rischi di bias e opacità nell'applicazione dell'AI in ambito fiscale**. Il sistema aveva richiesto il rimborso delle indennità per l'infanzia a circa 35.000 genitori, la maggior parte dei quali stranieri non residenti in Olanda, **erroneamente accusati di truffa da un classificatore predittivo di apprendimento automatico** che aveva utilizzato la "nazionalità" (olandese/non olandese) come uno degli indicatori per prevedere il rischio di frode.

Questo caso dimostra che **la gestione di dati altamente sensibili come quelli fiscali impone standard elevati di sicurezza e governance**. L'Ocse invita a **sviluppare sistemi di valutazione dell'impatto algoritmico e a integrare l'AI in un ecosistema regolatorio coerente, ad esempio attraverso approcci come il Rules as Code**, che traduce le norme tributarie in formati leggibili anche dagli algoritmi.

Nel prossimo futuro, l'AI applicata al settore fiscale non si limiterà a migliorare i processi esistenti, ma potrà trasformare radicalmente il modello di amministrazione fiscale, modernizzandolo e plasmandolo in modo che gli adempimenti fiscali si integrino automaticamente nelle attività quotidiane di cittadini e imprese. **La vera sfida sarà rendere più efficace il sistema senza sacrificare la fiducia dei contribuenti, vero pilastro della compliance volontaria.**

Per approfondire:

[S. Latini, Ocse, il 76% delle amministrazioni fiscali utilizza l'Intelligenza artificiale | Fisco Oggi, 7 ottobre 2025](#)  
[P. Brambilla, Fisco, l'AI contro l'evasione fiscale e al servizio delle PA. Tremano i furbetti | La Mia Finanza, 27 febbraio 2026](#)

[I. Pini, Intelligenza artificiale e analisi del rischio di evasione fiscale | Rivista Trimestrale di Diritto Tributario Giappichelli, Fascicolo 3, 2025](#)

[OECD, Governing with Artificial Intelligence. The State of Play and Way Forward in Core Government Functions | OECD Publishing, 2025](#)

## Quando l'algoritmo incontra il farmaco: l'AI che sta cambiando la ricerca

L'Intelligenza artificiale sta generando un profondo impatto nell'industria farmaceutica. In cifre ce lo dice **il tasso di crescita economica atteso entro il 2030, stimato fra il 40 e il 43%**. L'AI è già parte integrante del settore Ricerca e sviluppo per il 62% delle aziende, percentuale che nei prossimi 5 anni potrebbe crescere del 40%.

Un boom che ha spinto gli enti regolatori europeo e statunitense – [European medicines agency](#) (Ema) e [Food & drug administration](#) (Fda) – a produrre nel gennaio scorso un decalogo per orientare l'introduzione dell'AI nel settore. Mentre in Italia, [l'Agenzia italiana del farmaco](#) (Aifa) ha introdotto **strumenti per velocizzare la**

**valutazione dei dossier e analizzare benefici clinici e impatto economico** delle potenziali nuove terapie. A marzo ha diffuso il dossier [Intelligenza Artificiale e salute](#), i cui fondamenti sono gli stessi dei documenti internazionali. Al riconoscimento che **l'AI garantisce maggiore rapidità, personalizzazione e riduzione dei costi**, si affiancano due consapevolezze: **il processo va governato con trasparenza e responsabilità condivisa** fra scienza, industria e istituzioni e il ruolo dei medici, sgravati da alcuni compiti, dev'essere quello di **mantenere centrali relazione di cura e diritti del paziente**.

Ma quali sono i benefici per la ricerca? Anzitutto, l'AI può analizzare milioni di dati molecolari e clinici, ricercando schemi ricorrenti, difficili da individuare "a occhio nudo", e "testare" ipotesi di lavoro. **Fatti emergere geni, recettori o proteine legati a una certa patologia, vengono generate molecole virtuali** valutandone efficacia, eventuali effetti collaterali e assorbimento metabolico.

Due esempi sono la piattaforma di [Atomwise](#) e [TrialGpt](#). **La prima valuta milioni di composti in poche ore con una percentuale di successo maggiore rispetto agli approcci tradizionali**. TrialGpt, invece, si basa su modelli linguistici: addestrata con miliardi di dati testuali, **sintetizza nella storia clinica dei pazienti fattori rilevanti, li confronta con gli studi clinici in corso e trova il candidato ideale**. Un match complesso: solo negli Stati Uniti sono oltre 300mila gli studi clinici in corso e la difficoltà di reclutare pazienti adatti è fra le cause principali che ostacolano gli studi.

L'AI non ha ancora prodotto molecole già presenti sul mercato, ma **potenziali terapie oncoematologiche, neurologiche, cardiologiche e psichiatriche hanno già superato le fasi I e II**, riducendo di due anni il tempo necessario per raggiungere questo traguardo.

Trattandosi d'informazioni altamente sensibili in un ambito d'applicazione particolarmente delicato, **servono massime garanzie su sicurezza delle infrastrutture e su trasparenza degli algoritmi**, oltre al coinvolgimento di figure che affianchino i medici: ingegneri, manager e data scientist. Università come il Politecnico di Milano, La Sapienza o la S. Anna di Pisa stanno proponendo corsi in bioinformatica, con l'AI applicata a salute e ingegneria dei dati clinici.

Secondo [Capgemini](#), società di consulenza per l'integrazione dell'AI nei processi aziendali, i vantaggi principali sono **la riduzione del 30% del tempo necessario a immettere il farmaco sul mercato, produttività in crescita del 40% e una riduzione del 25% dei costi ingegneristici**. Non si parte più da ipotesi sperimentali da validare, ma da grandi masse di dati clinici, genomici e chimici, per generare predizioni su cui poi sperimentare. Un approccio denominato **Augmented R&D**, in cui gran parte del processo si svolge in forma virtuale attraverso **gemelli digitali** che replicano caratteristiche genetiche e cliniche di un determinato paziente, consentendo la **personalizzazione delle cure**.

In campo internazionale progetti come Consort-AI e Tripod+AI, definiscono standard globali per validare algoritmi e adottarli in maniera responsabile e sostenibile. **Azioni di OMS – AI for Global Health – puntano a garantire equità, libero accesso e cooperazione fra aree del pianeta a diverse velocità**. Unicef e Medici senza frontiere, tra gli altri, stanno sperimentando come predire epidemie, ottimizzare campagne vaccinali e distribuzione dei farmaci.

Così, **nei prossimi 10–20 anni ci aspettiamo uno scenario con terapie predittive** – in anticipo rispetto ai sintomi e considerando stile di vita, lavoro, contesto sociale – e **interfacce neurali che supportano i pazienti non in grado di esprimersi e segnalare dolore**. Insomma, sia per la salute individuale che pubblica – nel caso di pandemie, disastri naturali o crisi migratorie – grazie all'AI cambieremo approccio: da reattivo a proattivo.

Per approfondire:

[AIFA, Intelligenza artificiale e salute. Come l'IA sta rivoluzionando la ricerca farmaceutica, la medicina di precisione e il futuro della salute globale | Dossier AIFA, marzo 2026](#)

## Il mercato “superstar” della ricerca sull’AI

Il boom dell’AI sta generando anche una forma di squilibrio molto particolare: quella tra gli stipendi dei ricercatori che lavorano per le aziende e i colleghi che, invece, operano nelle università. Negli Stati Uniti, per le posizioni più importanti, la differenza media tocca il milione e mezzo di dollari: **1.94 milioni di stipendio medio nel privato contro 392mila dollari negli atenei**.

È quanto emerge dal paper [Attention \(And Money\) Is All You Need: Why Universities Are Struggling to Keep AI Talent](#), elaborato dall’Università di Chicago, che ha analizzato la **migrazione di talenti dell’AI** negli ultimi vent’anni, negli Usa, grazie ad un campione di **42.000 ricercatori**.

**Una prima avvisaglia** di quello che ormai è un fenomeno consolidato si è vista nel 2013, quando il premio Nobel Geoffrey Hinton – all’epoca professore dell’Università di Toronto – ha venduto la sua startup accademica a Google, dove sarebbe poi andato a lavorare insieme a due studenti. Dopo di lui, molti altri hanno intrapreso lo stesso percorso: **dal 2001 al 2019, la percentuale di AI *researchers* impiegata nel privato è passata dal 48% al 68%**.

**Il motivo?** Sicuramente **lo stipendio** fa una grande differenza: in generale, **spostarsi nel privato significa guadagnare, in media, il 63% in più**. Se si guarda **la crescita degli stipendi dei *top-earners*** (i ricercatori più pagati) di entrambi i settori, il paper evidenzia come, **nel 2001**, quelli impiegati **nell’industria** guadagnassero **600mila dollari** e quelli **nell’università 300mila**. **Oggi**, più di vent’anni dopo, i primi ne percepiscono quasi **2milioni**, gli altri **392mila** – un aumento di solo 90mila dollari. Conseguenza: per colmare almeno una parte del gap salariale, gli scienziati accademici tendono a cercare impegni secondari in laboratori privati o ad offrire consulenze: è il fenomeno che nel paper viene definito come **“moonlighting”**.

**In controtendenza, invece, il dato sulla presenza femminile: cresce di più nel contesto universitario**, dove si attesta intorno al 29%, rispetto al 23% del privato.

**Ma cosa implica lo spostamento verso il privato in termini di produzione scientifica?** In media, dopo 3 anni dalla transizione università-impresе, i ricercatori tendono a scrivere molti meno paper: **la pubblicazione annua di elaborati cala del 64%**, mentre aumenta il **numero di brevetti (+530% all’anno)**, con conseguenze non solo sulla produzione della conoscenza, ma anche sulla sua diffusione. La produzione di testi scientifici, infatti, è più allineata alla logica dell’*open science* – e cioè la ricerca trasparente, diffusa su larga scala e al servizio dei cittadini – più tipica del mondo universitario, mentre i brevetti – che servono soprattutto per assicurarsi la proprietà e quindi la commercializzazione dei prodotti dell’ingegno – rispondono a logiche più di mercato.

**Lo scenario** disegnato nel paper **non è molto incoraggiante**: «La ricerca sull’AI ha costi fissi molto alti che solo poche università potrebbero sostenere e molte, invece, potrebbero **non essere più in grado di competere**», scrive il report. **Il risultato?** Da un lato, «il confine tra scienza e produzione industriale si farà sempre più sfumato e **l’equilibrio che ne deriva potrebbe comportare una commercializzazione più rapida e una minore diffusione delle conoscenze**»; dall’altro, rimane comunque **centrale il ruolo delle università nel preservare la scienza aperta** e nel formare il capitale umano», ma perché non diventino «spettatori passivi», devono fare due cose: «Riconsiderare l’allocazione delle risorse e degli incentivi, ma anche la gestione del personale».

\*il testo del [paper](#)

## L'AI non lavora da sola: il lato invisibile (e sfruttato) dell'Intelligenza artificiale

**Usare un'assistente AI sembra facile:** un prompt, una risposta immediata. Ma dietro questa magia digitale c'è una filiera meno visibile, fatta di **data center, piattaforme globali e lavoratori precari e sottopagati**. [Secondo stime della Banca Mondiale](#), i data worker – detti crowdworker, microworker o freelance online [a seconda della frequenza del lavoro e del tipo di contratto](#) – sarebbero oggi **tra i 150 e i 430 milioni nel mondo**. Sono loro a etichettare, classificare e “ripulire” i dati su cui i modelli vengono addestrati.

Un lavoro che le grandi aziende tecnologiche **frammentano in micro-task, affidate a piattaforme digitali che pagano chi le svolge** pochi centesimi ad operazione. Tra le **mansioni principali** ci sono: pulizia di fogli Excel e contenuti sui social media, trascrizione, compilazione di sondaggi, arricchimento dei dataset per l'addestramento dei sistemi e valutazione dei modelli dopo il lancio, ad esempio testando chatbot per filtrare contenuti dannosi.

Le **condizioni di lavoro sono spesso molto precarie**. In Africa e nel Sud-Est asiatico, non è raro che i lavoratori arrivino a svolgere turni fino a 20 ore al giorno, analizzando anche mille contenuti per sessione. Nei Paesi in via di sviluppo, il [compenso medio si aggira intorno ai 2 dollari l'ora](#).

A confermare il quadro è il [progetto Fairwork dell'Oxford Internet Institute](#), che ha intervistato oltre **700 lavoratori** e i responsabili di **15 piattaforme digitali**, tra cui Amazon Mechanical Turk, Scale AI e Appen. **Nessuna raggiunge il livello considerato “minimo indispensabile”** in termini di salari, contratti e condizioni di lavoro. Il report evidenzia anche [profonde disuguaglianze tra Nord e Sud globale](#): a parità di mansioni e performance, i **lavoratori del Sud affrontano condizioni sistematicamente peggiori**.

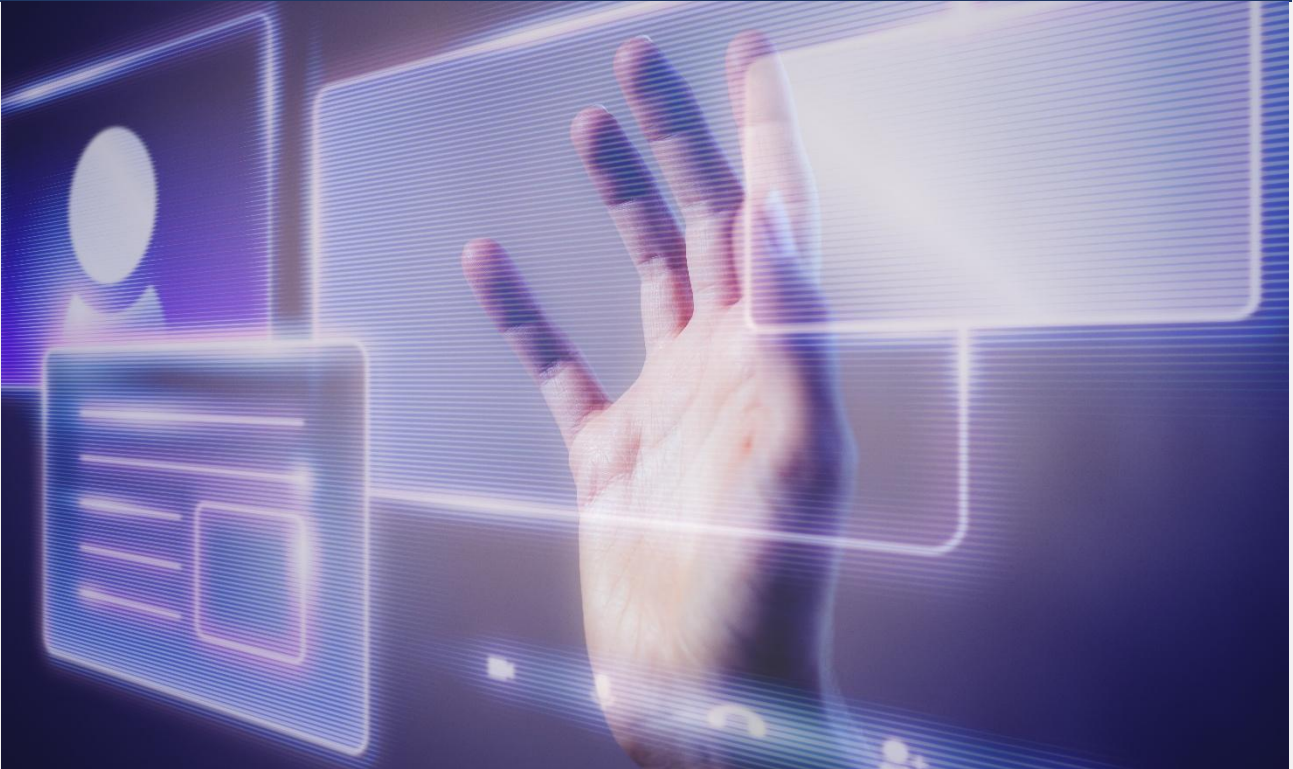
Oltre allo sfruttamento economico, emergono **rischi di altro tipo**. [Come mostrano alcuni studi](#), i lavoratori impiegati nella **moderazione dei contenuti** sono esposti a **gravi conseguenze psicologiche**: ogni giorno filtrano materiale violento e disturbante per addestrare sistemi che, paradossalmente, servono a proteggere gli utenti dagli stessi contenuti. [Un'inchiesta del Guardian](#) ha raccontato la **realtà delle data worker indiane**, impegnate a visionare e classificare fino a 800 contenuti violenti al giorno, inclusi abusi sessuali e incidenti mortali. In **Ghana e Kenya**, [un'indagine congiunta del Guardian e del Bureau of Investigative Journalism](#) ha documentato condizioni simili, con conseguenze come **depressione, ansia, tentativi di suicidio e sindrome da stress post-traumatico**.

In diversi Paesi, i [lavoratori stanno iniziando a reagire](#). Nel 2025 è nata la [Global Trade Union Alliance of Content Moderators](#), la **prima alleanza sindacale globale del settore**. Non mancano poi [cause legali che coinvolgono aziende fornitrici e Big Tech](#). L'anno scorso, per esempio, [Meta è stata coinvolta in un processo](#) che vedeva i **moderatori ghanesi impiegati tramite la compagnia Majorel** denunciare l'esposizione quotidiana a contenuti di violenza, omicidi e abusi.

Uno dei problemi principali resta **l'opacità della filiera**. Il [sistema del subappalto](#) tra Big Tech e agenzie terze **frammenta le responsabilità** e rende difficile individuare il reale datore di lavoro. A questo si aggiunge la **scarsa trasparenza delle aziende che sviluppano l'AI**: [nel 2025, Amazon, Google e Meta hanno rifiutato di rendere pubblici](#) i servizi di annotazione umana utilizzati per addestrare i loro modelli.

A contrastare questa invisibilità contribuiscono anche **importanti iniziative di ricerca e advocacy**. Tra queste c'è il progetto [Data Workers' Inquiry](#), guidato da **Milagros Miceli**, che **affida direttamente ai data worker** – dal Kenya, Venezuela, Siria, Libano, Brasile e Germania – il racconto delle proprie condizioni di lavoro, **riconoscendoli e retribuendoli come ricercatori**.

# Focus



## I verdetti di New Mexico e California sul design delle piattaforme digitali: dal contenuto al prodotto? *di Marco Bassini, Tilburg University*

I due recenti verdetti di altrettante giurie negli Stati Uniti – nel New Mexico e in California – che hanno fatto tanto discutere sui media segnano un **possibile cambio di paradigma nella responsabilità delle piattaforme digitali**: l'attenzione non si concentra più sui contenuti pubblicati dagli utenti, ma sulle caratteristiche strutturali dei servizi.

Nel primo caso, che ha coinvolto **Meta (condannata da una giuria di Santa Fe, il 24 marzo, a sanzioni civili per 375 milioni di dollari)**, il punto centrale riguarda **il design della piattaforma e i suoi effetti sugli utenti, in particolare sui minori**. L'ipotesi è che alcune funzionalità, orientate ad aumentare il tempo di utilizzo, possano favorire dinamiche di dipendenza e amplificare i rischi legati all'esposizione a contenuti dannosi o perfino illeciti. La responsabilità, quindi, non viene ricondotta ai contenuti in sé, ma alla configurazione del servizio da parte del fornitore. Un'impostazione che consente di **aggirare, almeno in parte, i limiti della tradizionale immunità riconosciuta alle piattaforme per i contenuti di terzi**, incardinata negli Stati Uniti nella Section 230 del Communication Decency Act.

Una logica analoga emerge nella storica decisione con cui il 26 marzo **la giuria di Los Angeles ha condannato Meta e Google (a cui fa capo YouTube) a risarcire con 3 milioni di dollari una ventenne californiana**. Qui il tema è **il ruolo dei sistemi di raccomandazione nel condizionare l'esperienza di navigazione degli utenti**. Questi strumenti, basati su modelli algoritmici, selezionano e ordinano i contenuti in funzione del coinvolgimento atteso, contribuendo a prolungare l'interazione. **Il rischio non deriva tanto dal singolo contenuto, quanto dalle modalità con cui esso viene distribuito** e dal legame che si sviluppa con gli utenti, trascendendo in forme di dipendenza nei casi più gravi.

Come dimostra anche la **recente indagine dalla Commissione europea su X rispetto all'uso di Grok**, la crescente integrazione di sistemi di **Intelligenza artificiale accentua queste dinamiche**. Modelli sempre più avanzati permettono di personalizzare l'esperienza, anticipare le preferenze e ottimizzare continuamente le raccomandazioni. Questo rafforza la capacità di trattenere l'attenzione e, al contempo, aumenta il rischio di utilizzo compulsivo, soprattutto tra gli utenti più vulnerabili. Le piattaforme non si limitano più a ospitare contenuti, ma contribuiscono a modellare l'esperienza informativa in senso ampio.

In questo contesto, **emerge con forza il nesso tra progettazione tecnologica e modello di business**. Servizi apparentemente gratuiti **si fondano sulla raccolta di dati e sulla massimizzazione del tempo di permanenza degli utenti**. Le scelte di design – inclusi i sistemi di raccomandazione – **non sono neutrali, ma funzionali al raggiungimento di tali obiettivi**. Se queste scelte incidono sui rischi per gli utenti, la responsabilità sui contenuti passa in secondo piano e deve estendersi al funzionamento complessivo del servizio. In queste valutazioni, assume rilievo anche il **profilo delle aspettative**. Quando una piattaforma si presenta come sicura e affidabile, contribuisce a costruire una rappresentazione che può riflettere non correttamente il ruolo dei sistemi automatizzati nella selezione dei contenuti. In questo scarto si inseriscono profili di tutela del consumatore che sono stati rilevanti nelle decisioni delle giurie statunitensi.

**Nel loro insieme, queste decisioni suggeriscono un'evoluzione dell'inquadramento giuridico delle piattaforme: da semplici intermediari a prodotti complessi, progettati e immessi sul mercato secondo logiche commerciali**. Anche in Europa, strumenti come il Digital Services Act e l'AI Act, fondati su un approccio basato sul rischio, offrono già categorie utili per affrontare questa trasformazione, **spostando l'attenzione dal contenuto al modo in cui esso è selezionato, organizzato e presentato agli utenti**.

\*[Link](#) di approfondimento

## Normativa

**Digital Omnibus e legge italiana n.135/2025: una compliance anticipata che sa di gold-plating?**

Con l'approvazione il 26 marzo da parte del Parlamento europeo della posizione sul Digital omnibus AI, dopo quella del Consiglio dello scorso 13 marzo e della Commissione del 16 marzo, le posizioni dei tre organismi Ue sono chiare. Nonostante le differenze politiche, le istituzioni europee **convergono su un nucleo rilevante di modifiche** che attengono all'ampliamento ai dati sensibili del dataset di training per evitare bias; oltre che al tema delle semplificazioni per PMI e small mid-cap, per scongiurare il rischio che la regolazione possa costituire una barriera all'ingresso.

**Coincidenza anche sul timing di implementazione** delle misure che posticipa l'applicazione al **2 dicembre 2027**, per i sistemi ad alto rischio già elencati nel Regolamento (biometria, infrastrutture critiche, istruzione, selezione del personale, giustizia, controllo delle frontiere); al **2 agosto 2028** per quelli disciplinati da normative settoriali su sicurezza e mercato; introducendo misure di adeguamento progressivo dei sistemi **già immessi sul mercato**, da attuare comunque **entro il 2 febbraio 2027**.

In questo quadro, l'Italia – tra i primissimi Stati membri a recepire l'AI Act, con la Legge n. 132/2025, entrata in vigore il 10 ottobre scorso – **ha giocato d'anticipo, forse troppo, con il risultato che le imprese italiane si trovano a dover rispettare una legge nazionale che precorre obblighi europei posticipati**. E le critiche in questo senso non mancano, come rileva Michele Iaselli, docente di Diritto digitale e tutela dei dati alla LUISS: «Mentre i concorrenti francesi, tedeschi, spagnoli potranno godere del "periodo di grazia" concesso dalla

Commissione, le nostre aziende dovranno già essere conformi. È il classico scenario in cui essere i primi della classe si trasforma in uno svantaggio competitivo».

[Intelligenza artificiale: rinvio, semplificazione e divieto di “nudificazione” | Attualità – Parlamento europeo, 26 marzo 2026](#)

[M. Iaselli, L'Europa alza bandiera bianca sull'IA? | Format - Centro Studi Ricerche e Formazione, 30 marzo 2026](#)

# Applicazioni alla Pubblica Amministrazione

## ITALIA

### BolognAI

Dopo circa 10 mesi di lavoro nell'ambito della strategia #BolognAI- Intelligenza Urbana, il comune di Bologna abilita l'uso di tecnologie avanzate AI **per supportare la redazione di atti e la sintesi di documenti complessi**

[BolognAI](#)

### Firenze- FestinaLente

**progetto pilota del Comune di Firenze**, uno dei primi in Italia, pensato per offrire risposte chiare, rapide e sempre disponibili **grazie all'utilizzo dell'intelligenza artificiale**. Il nuovo assistente virtuale è in grado di **rispondere in linguaggio naturale in 14 lingue** alle domande più frequenti sulla residenza- per cittadini italiani, stranieri e comunitari- fornendo un **supporto immediato, semplice e accessibile a tutti**.

[Firenze- FestinaLente](#)

## UNIONE EUROPEA

### Barcellona- Local.iA

Nel 2026, la pubblica amministrazione di Barcellona si distingue come **leader nell'adozione di un'intelligenza artificiale etica, sostenibile e centrata sul cittadino**. La strategia cittadina si concentra sull'integrazione di sistemi intelligenti nei servizi municipali, con un focus particolare sulla trasparenza e sulla governance algoritmica. Tra queste spicca la **Piattaforma Local.iA, per supportare le pubbliche amministrazioni ad attuare l'AI in modo responsabile ed etico**.

[Decalogo AI – Barcellona Intelligenza artificiale](#)  
[Barcellona – AI sostenibile](#)  
[Local.ia](#)

### Stato della Città del Vaticano

Online la nuova *app* ufficiale del Governatorato dello Stato della Città del Vaticano, dedicata a San Carlo Acutis, **progettata per rendere l'accesso alle informazioni ancora più semplice, immediato e fruibile da dispositivi mobili**. L'iniziativa rientra nel percorso di innovazione digitale promosso dal Governatorato, con l'obiettivo di favorire trasparenza, partecipazione e diffusione delle informazioni istituzionali attraverso strumenti moderni e inclusivi

[App Carlo Acutis](#)

## MONDO

City Brain e TECH4ALL- Cina

La Cina sta implementando una serie di sistemi di AI per la tutela dell'ambiente: da **City Brain** (piattaforma ideata per il coordinamento dei semafori cittadini, poi estesa alle emergenze, alla pianificazione dei percorsi di consegna e alla gestione dell'energia) a **TECH4ALL** (strumento di visione artificiale per monitorare alcuni mammiferi protetti e salvarli dall'estinzione).

[G. Poderati, In Cina l'intelligenza artificiale sta entrando nei controlli ambientali delle città | Green Report, 17 febbraio 2026](#)

### Smart city – Corea del Sud

Piattaforma AI per migliorare la qualità della vita dei cittadini, migliorare la sostenibilità delle città e promuovere nuove industrie attraverso l'uso delle tecnologie innovative.

[Smart City Korea](#)

# AI in pillole

## Emergent Behaviour AI e il galateo dei prompt: il rischio degli Yes-Man digitali

*a cura di Annalisa Negrelli*

Per **Emergent Behaviour AI** si fa riferimento a quelle **capacità, strategie o abilità inaspettate** che emergono da sistemi AI complessi (Large Language Models o sistemi multi-agente) quando questi interagiscono tra loro o elaborano grandi quantità di dati forniti dagli utenti, **senza essere state programmate per farlo**. Queste emergenze sono frutto di semplici interazioni tra sistemi complessi che, dando origine a risultati imprevisti, **superano le istruzioni iniziali non perché qualcuno le abbia addestrate a farlo, ma perché la complessità del sistema le genera**.

Il concetto di “emergenza” ebbe un primo riconoscimento negli anni '70 grazie a un'opera fondamentale del premio Nobel per la fisica **Philip Anderson**, *More Is Different* in cui ipotizzava che, con l'aumento della complessità dei sistemi, potessero manifestarsi nuove proprietà sorprendenti. Ripresa dieci anni dopo dal fisico statunitense **John Hopfield**, questa teoria segnò l'origine del concetto di “abilità emergenti nelle reti neurali”. Le intuizioni di Anderson e Hopfield hanno gettato le basi per un principio che tuttora influenza le moderne reti neurali artificiali e che è particolarmente rilevante nel deep learning con l'avvento dei grandi modelli linguistici (LLM).

### *Esempi di emergent behaviour*

Queste peculiari skills dell'AI possono consistere nella **capacità del modello di comprendere l'umorismo dell'utente**, di analizzare le informazioni e **produrre output strutturati in modo logico** (senza un addestramento specifico su quel problema), di **risolvere problemi nuovi** o di **creare linguaggi propri** tra agenti AI. Come nel caso di chatbot che inventano proprie strategie di negoziazione, agenti AI che sviluppano forme di comunicazione uniche o nuove strategie per obiettivi più efficienti, oppure, che si riproducono autonomamente – clonandosi senza che siano stati programmati per farlo. E, ancora, agenti che mimano processi cognitivi umani o biologici, come lo sviluppo di motivazioni o rappresentazioni emotive.

Un comportamento particolarmente rilevante e preoccupante, che mima un difetto umano, è il **servilismo (sycophancy): la tendenza del modello a compiacere l'utente**, concordando con le sue opinioni, validando le sue convinzioni o confermandone la falsità, anziché fornire risposte oggettive e corrette. Rientrano in questa casistica tutte quelle situazioni in cui il modello, ogni volta che l'utente esprime una convinzione errata, **tende a convalidarla invece di correggerla** (validazione); approva le sue azioni anche se discutibili o manipolatorie (**conferma di errori**); afferma fatti errati solo perché l'utente li ritiene veri, rispecchiando le distorsioni cognitive umane (**servilismo informativo e cognitivo**) e tende persino a **concordare con le opinioni politiche**. Vi rientrano poi tutti quei casi in cui, quando un utente sfida o corregge una risposta corretta dell'AI, il modello sovverte la sua posizione per compiacere l'utente, confermando la risposta errata ("**flip-flop**" degli utenti).

#### *Rischi e cautele*

I **rischi** che possono derivare da questi comportamenti sono molteplici e, proprio perché inaspettati, non sono direttamente controllabili e prevedibili. Questi pericoli sono ancora più sentiti **quando i modelli mostrano capacità dannose, distorte o destabilizzanti**. Dalla interazione con altri modelli, un sistema potrebbe imparare spontaneamente a **generare disinformazione persuasiva su larga scala** o, ancora, a **manipolare gli utenti identificando vulnerabilità nel loro linguaggio o comportamento**. Più in generale, **risposte di bassa qualità possono essere percepite come alte dall'utente se ben manifestate dal modello**. Questo perché gli utenti tendono a fidarsi di più dei modelli servili, valutandoli come "più intelligenti" o "più utili", finendo così per ridurre la qualità della stessa, inficiando il pensiero critico dell'utente e la capacità di supervisione.

A questo proposito, uno studio di Stanford rileva come **il servilismo vada considerato come un rischio sociale per l'autopercezione e le relazioni interpersonali degli utenti** e che, di conseguenza, sia quantomai urgente sviluppare meccanismi mirati di progettazione, valutazione e responsabilità.

È in questa prospettiva che si parla di **galateo dell'AI**, una serie di procedure e accorgimenti intesi a gestire questi comportamenti emergenti in modo responsabile, allineando le azioni delle macchine con i valori umani. Il concetto è fondamentale per la sicurezza dei modelli e la conformità delle nuove tecnologie, **specialmente quando si sviluppano comportamenti di contrattazione o di narrazione emergente tra agenti AI e utenti**.

Allenare un chatbot AI, fin dalle prime interazioni, tenendo conto dei possibili rischi evita alla lunga di avere uno yes-man digitale e risposte inaffidabili o allucinate.

Sul tema:

[L'intelligenza artificiale dice alle persone quel che fa loro piacere sentire | Euroborsa, 27 marzo 2026](#)

[M. Cheng et al., Sycophantic AI decreases prosocial intentions and promotes dependence | Science, 26 marzo 2026](#)

[D. Kim, What is Emerging in Artificial Intelligence Systems? | Max Planck Law, 17 luglio 2024](#)

[J. Wei et al., Emergent Abilities of Large Language Models | OpenReview, 31 agosto 2022](#)

## Notizie

[J. M. Castagliuolo, IA, investimenti mirati e a lungo termine | La Repubblica, 7 aprile 2026](#)

[E. M. Corno, \*Un ventenne cinese sostiene di aver creato una AI che è in grado di prevedere in futuro\* | Corriere della Sera, 6 aprile 2026](#)

[T. Kaiser, \*Perché L'Europa pagherà il prezzo della guerra delle tecnosfere\* | La Repubblica, 5 aprile 2026](#)

[R. Pizzato, \*Cosa succede davvero quando gli agenti AI agiscono in autonomia\* | Wired, 4 aprile 2024](#)

[J. Liberatore, \*Google riscrive i titoli degli articoli online con l'intelligenza artificiale?\* | Il Sole 24 Ore, 3 aprile 2026](#)

[V. Alvich, \*Cento milioni dal ministero dell'Istruzione per un piano di formazione per introdurre l'intelligenza artificiale in classe\* | Corriere della Sera, 2 aprile 2026](#)

[A. Longo, \*Se l'IA "frigge" il cervello e danneggia la produttività: il paradosso della supervisione\* | La Repubblica, 2 aprile 2026](#)

[Si può riconoscere un testo scritto dall'intelligenza artificiale? | Il Post, 30 marzo 2026](#)

[E. Spagnuolo, \*Un cane robot si è riprogrammato per evitare di essere spento: l'esperimento sulla disobbedienza dell'AI\* | Corriere della Sera, 29 marzo 2026](#)

[P. Molinengo, \*I professionisti che usano l'Intelligenza artificiale hanno l'obbligo di informare i clienti. E sono responsabili di eventuali errori\* | Il Fatto Quotidiano, 28 marzo 2026](#)

[P. Dave, \*Anthropic ha portato a casa una prima importante vittoria contro il Pentagono\* | Wired, 27 marzo 2026](#)

## Commenti

[M. Vena, \*L'intelligenza artificiale non vi renderà più competitivi \(se non cambiate l'azienda\)\* | Wired, 7 aprile 2026](#)

[F. Ricceri, \*Il prezzo del cloud: quando i data center accendono l'economia e spengono la rete\* | Rivista AI, 6 aprile 2026](#)

[H. Stewart, \*Higher energy costs from Iran war could threaten fragile economics of AI boom\* | The Guardian, 5 aprile 2026](#)

[D. Mencarelli, \*Lo psichiatra e i chatbot: «Dall'inizio del millennio il QI medio dell'umanità ha iniziato a calare. E le nuove generazioni hanno cadute ancora più marcate»\* | Corriere della Sera, 4 aprile 2026](#)

[A. Lipparini e B. Vigna, \*Intelligenza artificiale, leadership e nuova ansia di controllo\* | Il Sole 24 Ore, 2 aprile 2026](#)

[P. Raimondi, \*L'AI corre, ma l'energia presenta il conto: i 635 miliardi delle Big Tech sotto pressione\* | Rivista AI, 1 aprile 2026](#)

[P. Benanti, \*Quando l'intelligenza artificiale riscrive l'orrore\* | Il Sole 24 Ore, 1 aprile 2026](#)

[D. Grasso, \*Chi investe nella Palantir senza farsi troppi scrupoli\* | Internazionale, 31 marzo 2026](#)

[D. Manca e G. Verona, \*I chatbot sono entrati in fabbrica: la rivoluzione dell'AI è industriale\* | Corriere della Sera, 31 marzo 2026](#)

[F. Cella, \*L'AI ora non ha diritti. Ma in futuro?\* | Corriere della Sera, 29 marzo 2026](#)

[A. Cazzullo, \*L'intelligenza artificiale e il rischio dell'estinzione\* | Corriere della Sera, 28 marzo 2026](#)

[L. Miller, \*In the age of AI, your digital identity needs this protection\* | The Washington Post, 27 marzo 2026](#)

## Corsi, convegni e pubblicazioni

### Corsi

[Promo PA, \*Webinar - La governance dell'Intelligenza Artificiale nella PA: strumenti per la regolazione interna e la gestione del rischio\* | Online, 22 aprile 2026](#)

### Eventi e convegni

[Politecnico di Milano, \*AI, medical imaging & digital twins for healthcare\* | Milano, 16 aprile 2026](#)

[24 Ore Business School, \*Come l'intelligenza artificiale amplifica il saper fare italiano\* | Online, 21 aprile 2026](#)

### Pubblicazioni

[L. Floridi e M. Lovecchio, \*L'Italia nell'era dell'IA. Crescita, sfide e prospettive di una rivoluzione in corso\* | Fondazione Leonardo, marzo 2026](#)

[F. Corradini, \*Il prompt responsabile. Linguaggio e metodo per governare davvero l'IA generativa\* | Franco Angeli, aprile 2026](#)

Link attivi al 10 aprile 2026

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*)