

ATTUALITÀ

L'algoritmo è trasparente. Ma dati utilizzati e contenuti generati lo sono?

6 Ottobre 2026

Jacopo Liguori, Partner e Responsabile IP, Technology & Privacy, Withers



Jacopo Liguori, Partner e Responsabile IP, Technology & Privacy, Withers

> **Jacopo Liguori**

Jacopo Liguori è responsabile del team di proprietà intellettuale, tech e privacy negli uffici di Milano e Padova. Assiste primarie aziende italiane e internazionali e clienti privati, attivi in diversi settori quali: life science e pharma, fashion & luxury, automotive, food and beverage, sport, elettronica e software, entertainment e multimedia, assicurazioni, finanza e arte. Presta assistenza nel campo della proprietà intellettuale, delle nuove tecnologie, della privacy e della protezione dei dati e nel campo delle startup ad alto valore tecnologico.

1. Introduzione

Il dibattito sulla trasparenza dell'intelligenza artificiale si è sviluppato a una velocità che ha colto impreparati molti operatori, imprese e, in parte, gli stessi regolatori. In poco tempo si è passati da una discussione quasi accademica sulla "spiegabilità" degli algoritmi a un insieme di norme vincolanti, con scadenze ravvicinate e obblighi precisi. All'inizio il dibattito si è concentrato soprattutto su un aspetto: far capire con quale logica un modello arriva a una decisione, "aprendo la scatola nera". È un tema legittimo, ma non sufficiente. Nelle applicazioni concrete dell'IA generativa che oggi si usano negli uffici legali, nelle funzioni di compliance, nel servizio clienti e nel risk management - anche e soprattutto nei settori regolamentati - la questione non è solo più capire come ragiona il modello. Il problema più concreto ed attuale è controllare che cosa entra nel sistema e che cosa ne esce.

Questo spostamento di prospettiva non è solo teorico, ma è un aspetto su cui le aziende sono chiamate dal regolatore ad agire con tempestività. Dal 2 agosto 2026 è applicabile l'articolo 50 dell'AI Act, che impone obblighi di trasparenza specifici a fornitori e operatori di determinati sistemi di intelligenza artificiale, incluse le soluzioni generative e interattive e i cosiddetti deepfake. A ridosso di questa scadenza, il 20 luglio 2026, la Commissione europea ha adottato delle linee guida sugli obblighi di trasparenza per i fornitori e gli operatori di sistemi di IA (vedi qui linee guida C(2026) 5054). Si tratta di un documento di oltre cinquanta pagine che definisce l'ambito di applicazione degli obblighi di trasparenza per i fornitori e gli operatori di sistemi di IA a norma dell'articolo 50 della legge sull'AI. Le linee guida non sono vincolanti - solo la Corte di Giustizia può fornire un'interpretazione autentica dell'AI Act - ma è ragionevole attendersi che le autorità nazionali di vigilanza del mercato e l'AI Office le seguano nella prassi applicativa, il che le rende, nei fatti, un parametro di riferimento ineludibile per chiunque debba dimostrare la propria conformità.

2. La falsa promessa della "scatola trasparente"

C'è un paradosso che vale la pena mettere subito in luce: anche un modello perfettamente spiegabile può produrre decisioni sbagliate, discriminatorie o dannose, se i dati che lo alimentano sono scorretti, incompleti o non pertinenti. È la versione aggiornata di un principio antico dell'informatica, il "garbage in, garbage out": se i dati in ingresso sono viziati, non validi o inaffidabili, i risultati generati non potran-

no essere attendibili perché nessuna trasparenza algoritmica potrà mai correggere l'output.

Le organizzazioni che investono energie soltanto nella spiegabilità tecnica del modello, trascurando la qualità e la tracciabilità di ciò che vi entra e di ciò che ne esce, rischiano di costruire una compliance apparente: formalmente corretta, sostanzialmente cieca rispetto ai rischi reali.

3. Trasparenza degli input: il grande assente del dibattito

Il primo fronte, ancora poco presidiato nella prassi aziendale, riguarda la provenienza e la qualità dei dati inseriti nei sistemi di IA. Si tratta di dati aziendali riservati, dati personali di dipendenti o clienti, dati pubblici raccolti da fonti eterogenee, oppure dati sintetici generati per addestrare o testare i modelli. Ognuna di queste categorie porta con sé un diverso profilo di rischio: dalla violazione di obblighi di riservatezza contrattuale, alla compliance con il GDPR, fino alla tutela del know-how e dei segreti industriali quando i contenuti caricati in un sistema di IA finiscono, magari inconsapevolmente, per alimentare l'addestramento di un modello di terzi.

C'è poi un secondo livello, ancora meno esplorato: la qualità dei prompt. Nel momento in cui un dipendente formula un'istruzione a un sistema di IA generativa, sta compiendo una scelta che orienta il risultato tanto quanto un processo decisionale aziendale tradizionale. Un prompt formulato in modo impreciso, parziale o orientato può introdurre bias, distorcere l'analisi o produrre conclusioni non affidabili, anche in presenza di un modello tecnicamente ineccepibile. In questo senso, il prompt diventa un vero e proprio atto organizzativo, che meriterebbe di essere trattato con lo stesso rigore riservato, ad esempio, all'approvazione di un contratto o di una policy interna.

4. Trasparenza degli output: informare non basta

Sul versante opposto, l'AI Act sposta l'attenzione sugli obblighi informativi verso l'utente finale. L'articolo 50, paragrafo 1, richiede che i fornitori progettino i sistemi destinati a interagire direttamente con le persone fisiche in modo tale che queste siano informate, non oltre la prima interazione, di stare interagendo direttamente con un sistema di intelligenza artificiale, salvo che ciò non risulti già evidente dal contesto. Il paragrafo 2 dello stesso articolo impone invece ai fornitori di sistemi che generano contenuti sintetici audio, video, immagini o testuali di marcare tali output in un formato leggibile da

macchina, così da renderli riconoscibili come artificialmente generati o manipolati.

Le linee guida della Commissione precisano diversi aspetti operativi che meritano attenzione. Il paragrafo 34 prevede infatti che l'informazione possa essere veicolata attraverso strumenti testuali, sonori, visivi o multimodali, purché sia resa in modo chiaro, distinguibile e adeguato alle caratteristiche degli utenti, in particolare di quelli vulnerabili.

Le informazioni nascoste nei termini e condizioni, o segnali ambigui e poco percepibili, non sono considerati sufficienti. Le linee guida raccomandano inoltre al paragrafo 39 di calibrare la disclosure sul livello di rischio del contesto: un avviso singolo e ben visibile può bastare in situazioni ordinarie, mentre contesti più delicati — ad esempio l'interazione con utenti vulnerabili o l'erogazione di consulenza con effetti economici rilevanti — possono richiedere promemoria periodici, evitando al contempo il fenomeno della cosiddetta *"banner blindness"*, ossia l'abitudine dell'utente a ignorare avvisi ripetuti e standardizzati.

Quanto agli obblighi di marcatura di cui all'articolo 50, paragrafo 2, il Regolamento (UE) 2026/1744 ha introdotto una disciplina transitoria per i sistemi di IA, inclusi i sistemi di IA per finalità generali, che generano contenuti sintetici e che erano già stati immessi sul mercato prima del 2 agosto 2026: i relativi fornitori devono adottare le misure necessarie per conformarsi all'articolo 50, paragrafo 2, entro il 2 dicembre 2026: un termine da tenere sotto controllo per chi utilizza soluzioni di IA generativa integrate in prodotti o servizi già esistenti.

Ma limitarsi a un avviso generico (*"stai parlando con un'IA"*) rischia comunque di essere un adempimento formale privo di sostanza, se non è accompagnato da informazioni su quanto il risultato sia stato verificato da una persona, su quali siano i limiti noti del sistema e su quale grado di affidabilità possa ragionevolmente attribuirsi all'output. La vera trasparenza informativa non si esaurisce nella dichiarazione iniziale, ma accompagna l'intero processo di utilizzo del contenuto generato.

5. Il banco di prova dei settori regolamentati

Il settore finanziario è probabilmente il terreno in cui questa tensione tra trasparenza degli input, trasparenza degli output e obblighi settoriali preesistenti si manifesta con maggiore intensità, perché alla

cornice orizzontale dell'AI Act si sovrappongono discipline già consolidate: dal Testo Unico Bancario al Regolamento DORA, fino alla normativa antiriciclaggio e, in Italia, alla legge 23 settembre 2025, n. 132, che affianca l'AI Act rafforzando il coordinamento tra Banca d'Italia, Consob e Ivass. Alcuni esempi concreti aiutano a rendere tangibile quanto fin qui argomentato in astratto.

- **Scoring creditizio e valutazione del merito di credito:** l'Allegato III, punto 5, lettera b), dell'AI Act classifica come ad alto rischio i sistemi utilizzati per valutare l'affidabilità creditizia delle persone fisiche. Qui la trasparenza degli input assume un significato molto concreto: quali dati alimentano il modello (dati reddituali, comportamentali, talvolta fonti alternative come i dati di pagamento o i social media), quanto sono aggiornati e rappresentativi, e se la loro selezione introduce discriminazioni indirette verso categorie di clienti. La banca, in qualità di titolare del trattamento e, ove applicabile, di deployer o provider ai sensi dell'AI Act, deve essere in grado di ricostruire il processo decisionale automatizzato sia per garantire l'esercizio dei diritti previsti dall'articolo 22 GDPR, sia per fornire le informazioni sulla logica utilizzata e sulle conseguenze del trattamento richieste dagli articoli 13, 14 e 15 GDPR oltre che agli obblighi di documentazione tecnica e supervisione umana già previsti dall'AI Act per i sistemi di AI ad alto rischio.
- **Antiriciclaggio e fraud detection:** i sistemi di transaction monitoring e di screening delle liste sanzionatorie sollevano un problema speculare. L'input critico non è tanto il dato del singolo cliente, quanto la qualità e l'aggiornamento delle liste e dei parametri di rischio utilizzati per generare gli alert. Un output che porti al blocco di un conto o al rifiuto di un'operazione può, a seconda del ruolo che l'AI assume nella decisione finale, rendere necessario verificare se il sistema rientri tra quelli ad alto rischio ai sensi dell'Allegato III; anche quando ciò non accade, resta comunque essenziale poter tracciare perché un determinato alert è stato generato e chi lo ha validato, prima di comunicarlo alle autorità competenti.
- **Chatbot, assistenti virtuali e robo-advisory:** l'articolo 50, paragrafo 1, si applica direttamente alle chatbots di customer service e agli assistenti virtuali sempre più diffusi negli sportelli digitali e nelle app bancarie. Ad esempio, una banca utilizza un assistente AI sul proprio sito per spiegare caratteristiche di conti correnti, fornire informazioni sui mutui, indirizzare il cliente verso un operatore. L'utente deve essere informato chiaramente che sta interagendo con un sistema

AI. Non basta scriverlo nelle condizioni generali ma serve una disclosure visibile all'inizio della conversazione. Le linee guida della Commissione chiariscono inoltre che, in contesti caratterizzati da un maggiore rischio di errore o affidamento da parte dell'utente, come la consulenza finanziaria automatizzata, la semplice disclosure iniziale potrebbe non bastare: sono raccomandati promemoria periodici e un linguaggio che renda evidente, in ogni fase dell'interazione, che non si sta parlando con un consulente umano.

- **Report di mercato generati da AI:** pensiamo infine ad un operatore che intenda pubblicare sul proprio sito un commento settimanale sui mercati finanziari generato da AI. Se il testo viene pubblicato automaticamente è probabile un obbligo di disclosure AI. Ma se il testo è controllato, corretto e validato da un analista e da un responsabile editoriale e invece probabile che si applichi l'eccezione per human review/editorial control.
- **Esternalizzazione a fornitori terzi e Regolamento DORA:** quando il sistema di IA è fornito in outsourcing da un soggetto terzo, la trasparenza degli input e degli output diventa anche un tema contrattuale. Le clausole di exit strategy, le previsioni su restituzione e cancellazione dei dati, gli obblighi di audit e le evidenze documentali richieste da DORA per i servizi ICT critici si intrecciano direttamente con la capacità dell'ente di dimostrare, in caso di ispezione, quali dati sono transitati nel sistema e con quali garanzie.

Questi esempi mostrano come, nei settori regolamentati, la trasparenza dell'IA non sia un tema isolato da affrontare in una policy dedicata, ma un elemento che si innesta in modo pervasivo negli obblighi già esistenti in materia di vigilanza prudenziale, antiriciclaggio, protezione dei dati e resilienza operativa digitale.

6. Dalla trasparenza alla tracciabilità

Il passaggio più rilevante, dal punto di vista della governance aziendale, è forse questo: la vera trasparenza non consiste nel rendere pubblico il funzionamento interno dell'algoritmo, quanto piuttosto nel poter ricostruire, in qualsiasi momento, il percorso che collega un input, il modello utilizzato e l'output prodotto. È un concetto molto vicino all'accountability che il GDPR ha reso familiare ai responsabili privacy: se una scelta non è documentata, non è dimostrabile; e se non è dimostrabile, in caso di con-

testazione, è come se non fosse mai stata presa.

Le domande che molte organizzazioni stanno iniziando a porsi sono in realtà molto pratiche: siamo in grado di sapere quali dati aziendali vengono caricati nei sistemi di IA utilizzati dai dipendenti? sappiamo documentare quando e come un contenuto generato dall'IA è stato validato da una persona? i contratti con i nostri fornitori di IA ci consentono di accedere alle informazioni necessarie in caso di audit, contestazioni o richieste delle autorità? In molti casi il problema non è tanto l'assenza di regole, quanto l'assenza di processi che consentano di dimostrare il rispetto delle regole già esistenti.

Applicato all'intelligenza artificiale, questo principio si traduce in pratiche operative concrete che le organizzazioni dovrebbero iniziare a strutturare fin da ora:

- tenere un log dei prompt rilevanti, soprattutto quando l'output alimenta decisioni con effetti su clienti, dipendenti o terzi;
- documentare la versione del modello utilizzato e gli eventuali aggiornamenti che ne modificano il comportamento;
- conservare gli output più significativi insieme al contesto in cui sono stati generati;
- tracciare chi ha validato il risultato, quando e con quale livello di revisione umana;
- definire, per le funzioni più esposte (legale, HR, comunicazione), un presidio che intervenga prima dell'adozione di un nuovo strumento di IA, non dopo che è già in uso.

Non si tratta di burocratizzare l'innovazione, ma di costruire quella che potremmo chiamare una "filiera della fiducia": una catena documentale che consenta, in caso di errore, contestazione o richiesta di un'autorità di controllo, di ricostruire con chiarezza che cosa è stato chiesto al sistema, come ha risposto e chi ha validato il risultato prima che producesse effetti concreti. Tradurre questi principi in un sistema di regole interne coerente, che tenga conto della normativa di settore applicabile e che regga alla prova di un'ispezione o di un contenzioso, richiede tuttavia scelte tecnico-giuridiche non banali, spesso da costruire caso per caso insieme a chi conosce a fondo sia l'AI Act sia la disciplina settoriale con cui questo si intreccia.

7. Un tema di governance, non solo di tecnologia

Le linee guida europee sull'articolo 50 dell'AI Act rappresentano dunque un tassello importante, ma non esauriscono il tema. La trasparenza dei sistemi di intelligenza artificiale è destinata a intrecciarsi sempre più con la disciplina della protezione dei dati personali, con la tutela della proprietà intellettuale e del know-how aziendale, e con le regole di risk management già note ai settori più regolamentati, dalla finanza alle infrastrutture critiche.

Le organizzazioni che oggi si limitano a inserire un banner "*generato con l'IA*" nei propri contenuti, senza interrogarsi su che cosa sia entrato nel sistema per generarli e su chi ne abbia verificato l'attendibilità, stanno affrontando solo metà del problema. Nel prossimo futuro, le aziende non saranno valutate soltanto per la qualità dei sistemi di intelligenza artificiale che hanno adottato, ma per la loro capacità di dimostrare da dove provengono i dati che li alimentano, come sono stati elaborati e perché i risultati che ne derivano possano considerarsi affidabili. La trasparenza dell'intelligenza artificiale, in altre parole, non è più soltanto una questione tecnologica: è diventata, a tutti gli effetti, una questione di governance.

Per le organizzazioni che operano in settori regolamentati, il momento di interrogarsi su questi aspetti non è quando arriva una richiesta ispettiva o una contestazione, ma prima: nella fase in cui si negoziano i contratti con i fornitori di IA, si definiscono le policy interne di utilizzo e si stabilisce chi, in azienda, deve validare che cosa. È un lavoro che richiede una lettura congiunta dell'AI Act, della normativa di settore e delle regole di protezione dei dati personali, per costruire presidi che reggano alla prova di un'ispezione o di un contenzioso.



DB non solo
diritto
bancario

A NEW DIGITAL EXPERIENCE

 **dirittobancario.it**
