Governare gli agenti di AI: cosa devono considerare le organizzazioni

Secondo il PwC 2025 AI Agent Survey, il 79% dei dirigenti senior intervistati ha confermato che gli agenti di AI sono già in fase di adozione all'interno delle proprie organizzazioni. Questi sistemi fanno molto più che supportare il processo decisionale. A differenza dei LLM o dei chatbot, gli agenti sono in grado di recuperare informazioni in modo autonomo, interagire con diversi sistemi e agire con un coinvolgimento umano limitato.

Questo livello di autonomia modifica il profilo di rischio e solleva nuove sfide in termini di governance, protezione dei dati e accountability. Il rischio principale per qualsiasi organizzazione non è l'utilizzo degli agenti di AI in sé, bensì la loro implementazione in assenza di framework adeguati, competenze specifiche e meccanismi di supervisione in grado di gestirne l'impatto.

In questo blog analizziamo cosa sono gli agenti di AI, come la regolamentazione dell'AI varia nelle diverse giurisdizioni e le principali considerazioni in materia di accountability e protezione dei dati che le organizzazioni dovrebbero affrontare prima della loro implementazione.

In particolare, esaminiamo:

- **Cos'è un agente di AI e perché modifica il profilo di rischio**
- **La regolamentazione dell'AI a livello globale e l'operatività degli agenti di AI in contesti transfrontalieri**
- **Suggerimenti per una governance solida degli agenti di AI**

Che cos'è un agente di AI e perché modifica il profilo di rischio?

Un agente di AI è un sistema progettato per operare con un certo grado di autonomia, andando oltre i semplici strumenti basati su prompt e risposta. In pratica, gli agenti di AI sono in grado di pianificare ed eseguire attività articolate in più fasi, formulare decisioni o raccomandazioni, recuperare informazioni da molteplici fonti di dati e interagire sia con sistemi interni sia con strumenti di terze parti.

Ciò che distingue gli agenti di AI dai più familiari chatbot e LLM è il fatto che operano all'interno dei sistemi, e non si limitano a generare output. Per questo motivo, governance, accountability e gestione degli accessi ai dati diventano elementi essenziali da considerare già in fase di progettazione.

Principali rischi di governance legati agli agenti di AI e come mitigarli

Gli agenti di AI possono presentare criticità specifiche in funzione del loro specifico utilizzo, ma alcuni rischi di governance sono comuni a tutte le implementazioni di agenti di AI, indipendentemente dal settore.

1. Gli accessi possono estendersi oltre quanto previsto

Nel tempo, è frequente che vengano concessi ulteriori accessi senza riesaminare la valutazione del rischio originaria. Se gli accessi non sono limitati e riesaminati regolarmente, il principio di minimizzazione dei dati può essere compromesso.

Suggerimento: *applicare il principio del least privilege e rivedere periodicamente le autorizzazioni, per assicurare che restino coerenti con la finalità definita dell'agente.*

2. Le decisioni dell'AI non hanno un unico responsabile chiaramente identificato

Quando un agente attiva azioni lungo più workflow, spesso sono coinvolti diversi dipartimenti. L'organizzazione nel suo complesso rimane responsabile, ma la titolarità delle decisioni risulta frammentata.

Suggerimento: *assegnare una responsabilità chiara per ciascun agente di AI, includendo compiti di supervisione, gestione delle criticità (procedure di escalation) e approvazione finale.*

3. Mancanza di tracciabilità nei processi decisionali multi-step

Quando gli agenti pianificano azioni articolate in più fasi, può risultare difficile dimostrare quali dati siano stati utilizzati e perché sia stato scelto un determinato percorso decisionale. Questo diventa particolarmente critico in sede di audit o nella gestione di una richiesta di accesso ai dati personali.

Suggerimento: *garantire che gli agenti di AI generino log di audit e registrazioni delle decisioni adeguate, in modo che le azioni possano essere spiegate quando necessario.*

4. Modifiche successive alla messa in esercizio

Aggiornamenti a prompt, modelli, fonti di dati o strumenti possono ampliare gradualmente le funzionalità di un agente, spesso senza una chiara rivalutazione dei rischi associati. Nel tempo, ciò può portare a una deriva funzionale (c.d. function creep) e a trattamenti non intenzionali.

Suggerimento: introdurre controlli formali sulle modifiche, prevedendo che ogni aggiornamento rilevante attivi una revisione dei rischi, delle misure di governance e delle approvazioni necessarie prima del rilascio in produzione.

5. Errori di piccola entità che si propagano immediatamente

Quando decisioni basate su AI incidono su clienti o dipendenti, anche errori minimi possono diffondersi rapidamente. Ciò aumenta il rischio regolatorio e può compromettere in tempi brevi la fiducia di clienti e stakeholder.

Suggerimento: *integrare misure di salvaguardia, come soglie di revisione umana ed efficaci meccanismi di escalation per le decisioni a rischio più elevato.*

La regolamentazione dell'AI a livello globale e l'operatività transfrontaliera degli agenti di AI

Una delle principali sfide per le organizzazioni è che la governance dell'AI non è standardizzata a livello internazionale. Per le imprese che operano in più giurisdizioni, ciò significa che la governance non può essere progettata facendo riferimento a un unico regime regolatorio.

Le organizzazioni hanno quindi bisogno di controlli di governance dell'AI in grado di scalare oltre i confini nazionali, adattarsi ad aspettative regolatorie differenti e soddisfare i requisiti più stringenti. Questo aspetto è particolarmente rilevante per gli agenti di AI, che spesso operano simultaneamente su più sistemi, dataset e aree geografiche.

Di seguito analizziamo come la governance dell'AI è disciplinata nelle diverse aree del mondo.

Unione Europea

L'UE ha adottato un approccio basato sul rischio attraverso l'AI Act, con obblighi calibrati sulle modalità di utilizzo dei sistemi di AI e sul livello di rischio che essi comportano per gli individui. Tale quadro si affianca al GDPR, che continua a disciplinare il trattamento dei dati personali.

Sebbene l'AI Act UE sia entrato in vigore nel 2024, la maggior parte dei suoi obblighi sarà applicabile

a partire dal 2026. Questo segna il passaggio dalla fase di elaborazione delle politiche a quella di applicazione attiva, sia a livello europeo sia nazionale, con un conseguente incremento dello scrutinio regolatorio.

Per le organizzazioni che implementano agenti di AI, ciò implica che la governance debba essere attentamente valutata e approfondita prima della messa in esercizio di tali agenti. La classificazione dell'agente, i contesti di utilizzo e il ruolo dell'organizzazione nel ciclo di vita dell'AI incidono tutti sul livello di supervisione, documentazione e controllo atteso.

Scopri di più sull'AI Act.

Regno Unito

Il Regno Unito non ha introdotto una normativa unica e specifica sull'AI. Si affida invece a un approccio basato su principi, supportato dalla legislazione esistente, con il UK GDPR che riveste un ruolo centrale.

Per gli agenti di AI, l'articolo 22 del UK GDPR rappresenta spesso un punto critico, in particolare quando i sistemi effettuano decisioni interamente automatizzate che producono effetti giuridici o analogamente significativi sugli individui. Le organizzazioni devono quindi essere certe di dove avvenga l'automazione, di come l'intervento umano rimanga garantito e di come le decisioni possano essere contestate.

Canada

Il Canada non dispone ancora di una normativa organica e specifica sull'AI in vigore. La proposta di legislazione federale contenuta nel Bill C-27, inclusa l'Artificial Intelligence and Data Act (AIDA), è attualmente sospesa, generando incertezza sugli obblighi futuri a livello nazionale.

Ciò non significa, tuttavia, che gli agenti di AI operino in un vuoto normativo. La governance dell'AI in Canada è attualmente modellata dal Personal Information Protection and Electronic Documents Act (PIPEDA), che disciplina il trattamento dei dati personali, e dai regimi provinciali, come la Law 25 del Québec.

Per le organizzazioni, questo comporta la necessità di adottare approcci di governance sufficientemente robusti da soddisfare gli obblighi esistenti, ma al contempo flessibili per adattarsi all'evoluzione normativa.

Stati Uniti

Gli Stati Uniti non dispongono attualmente di una legge federale unica sull'AI. La governance dell'AI sta invece emergendo attraverso un mosaico di normative statali, affiancate da leggi in materia di privacy, tutela dei consumatori e antidiscriminazione.

Nel dicembre 2025, il Presidente Trump ha emanato un **Executive Order** sull'AI con l'obiettivo di limitare un'ulteriore frammentazione normativa, scoraggiando divergenze a livello statale. L'ordine promuove un approccio nazionale unitario e, pur non introducendo una legge federale vincolante, utilizza il potere esecutivo per influenzare l'azione degli Stati e porre le basi per una futura legislazione.

Per le organizzazioni che implementano agenti di AI, questo scenario crea una fase di transizione, più che di certezza normativa. I requisiti statali continuano ad applicarsi, soprattutto in contesti ad alto rischio come l'occupazione, la finanza e il processo decisionale nei confronti dei consumatori.

Per le organizzazioni operative in più Stati, rimangono essenziali una governance scalabile e una documentazione riutilizzabile, sia per gestire gli obblighi statali attuali sia per adattarsi rapidamente all'eventuale sviluppo di un quadro nazionale.

Suggerimenti per una governance solida degli agenti di AI

- **Assegnare una responsabilità chiara**

Le organizzazioni dovrebbero stabilire chi è responsabile dell'approvazione dei casi d'uso degli agenti di AI, della validazione delle valutazioni del rischio e della supervisione delle prestazioni nel tempo e delle modifiche apportate. Tale responsabilità può essere attribuita a un AI Officer oppure rientrare in un ruolo già esistente, come quello del Data Protection Officer (DPO). Ciò che conta maggiormente non è la denominazione del ruolo, ma che l'accountability sia definita in modo chiaro.

AI Officer vs DPO

• Valutare i rischi fin dalle fasi iniziali

Prima della messa in esercizio, le organizzazioni dovrebbero valutare se sia necessario o opportuno condurre un AI Impact Assessment (AIIA), in particolare quando gli agenti operano in modo autonomo, su larga scala o incidono sugli esiti che riguardano le persone.

Gli AIIA dovrebbero essere utilizzati in combinazione con le Data Protection Impact Assessment (DPIA), offrendo una visione più ampia del comportamento del sistema, dei rischi e delle misure di salvaguardia.

AI Impact Assessments

• Definire confini chiari

È fondamentale stabilire limiti precisi rispetto a ciò per cui un agente di AI è progettato. Ciò include la documentazione della finalità dell'agente e dei risultati attesi, delle decisioni che può assumere o influenzare, delle azioni espressamente vietate e delle fonti di dati e degli strumenti a cui è autorizzato ad accedere.

Confini chiaramente definiti aiutano a prevenire la deriva funzionale (scope creep) e supportano il principio di minimizzazione dei dati.

• Garantire le competenze adeguate

Una governance efficace dell'AI dipende dalla disponibilità di un mix adeguato di competenze ed esperienza. Le organizzazioni dovrebbero assicurarsi di avere accesso a competenze tecniche, di prodotto, legali e di compliance sufficienti per progettare, valutare e supervisionare gli agenti di AI lungo l'intero ciclo di vita. Nella pratica, è spesso in questo ambito che emergono in modo più evidente i gap di competenze.

• Costruire una supervisione umana significativa

La supervisione umana dovrebbe essere proporzionata al livello di rischio ed essere in grado di consentire un intervento effettivo. Ciò può includere la validazione degli output, la definizione di meccanismi di segnalazione e gestione dei casi critici (c.d. escalation), la facoltà di revisione e modifica della decisione automatizzata (c.d. override) e il mantenimento di adeguate tracce di audit. Questo aspetto è particolarmente rilevante nei casi in cui trovino applicazione processi decisionali automatizzati.

• Monitorare bias ed esiti non equi

Le organizzazioni dovrebbero adottare misure per individuare e monitorare potenziali bias o risultati discriminatori introdotti dagli agenti di AI. Ciò include l'analisi degli output nel tempo, la valutazione dei pattern decisionali e la verifica che i processi automatizzati non producano svantaggi ingiustificati per individui o gruppi.

• Integrare la governance nei processi esistenti

La governance dell'AI non dovrebbe essere isolata. Gli agenti di AI dovrebbero essere integrati nei Registri delle attività di trattamento (RoPA), nei processi di gestione del rischio, nella documentazione privacy e nei framework di sicurezza esistenti.

La governance deve rimanere dinamica, con una revisione periodica dei rischi man mano che sistemi, casi d'uso e contesti normativi evolvono.

The DPO Centre collabora a stretto contatto con le organizzazioni per valutare i rischi legati all'AI e fornire un supporto pratico di governance per l'implementazione dei sistemi di AI in modo consapevole e sicuro.

Contattaci per saperne di più sui nostri servizi in materia di AI.