AI agentici e sicurezza delle identità: perché il vero rischio è l’eccesso di privilegi

Ogni delega di responsabilità a una macchina ridisegna il perimetro della fiducia aziendale, e con gli AI agentici quel perimetro non passa più dai confini della rete ma dalle credenziali che muoviamo al suo interno: è qui che la sicurezza delle identità smette di essere un presidio tecnico tra tanti e diventa la lente attraverso cui leggere ogni rischio, perché ciò che decide quanto possiamo fidarci degli agenti AI non è quello che sanno fare, ma quello a cui sono autorizzati ad accedere.

I team di sicurezza stanno ancora cercando di comprendere fino in fondo l’impatto che l’intelligenza artificiale sta avendo sul panorama delle minacce informatiche. Molti osservatori avevano previsto l’arrivo di una nuova generazione di rischi, tale da richiedere approcci completamente diversi alla difesa delle reti aziendali.

La realtà, però, è meno futuristica e più concreta: nella maggior parte dei casi, l’AI non crea minacce totalmente nuove, ma amplifica e accelera vulnerabilità già esistenti. E quando si parla di AI agentica, il punto di osservazione più utile è quello che prende in esame la sicurezza delle identità digitali.

Il nodo centrale è semplice: quando gli agenti AI ricevono privilegi eccessivi, diventano un insider risk potenzialmente più pericoloso di un dipendente negligente o malintenzionato.

Alcune best practice di sicurezza già esistenti possono certamente essere adattate a questo nuovo scenario, ma il

problema è che il panorama delle minacce evolve a una velocità superiore rispetto alla capacità delle organizzazioni di governarlo.

La nuova ondata dell’AI agentica

L’AI agentica rappresenta la nuova fase di una trasformazione tecnologica che sta cambiando rapidamente il modo in cui le aziende operano. Secondo McKinsey, il 23% delle organizzazioni sta già scalando soluzioni di AI agentica in almeno una funzione aziendale, mentre un ulteriore 39% sta sperimentando queste tecnologie.

Nel settore bancario, ad esempio, gli “agent squad” vengono già utilizzati per automatizzare processi di mutuo e compliance. Le telco li impiegano nel customer service. L’obiettivo è chiaro: aumentare l’efficienza operativa, ridurre errori e costi e migliorare l’esperienza dei clienti.

Tuttavia, i rischi associati a questa evoluzione sono ancora largamente sottovalutati.

A differenza dell’AI generativa tradizionale, che produce contenuti o sintetizza informazioni in risposta a un prompt umano, gli agenti AI operano in modo autonomo attraverso un ciclo cognitivo continuo: percepiscono, ragionano, agiscono e apprendono.

Gli agenti raccolgono dati da fonti differenti, li elaborano per comprenderne il contesto e sfruttano i large language model per prendere decisioni e stabilire priorità operative. Successivamente eseguono azioni interagendo con strumenti e sistemi esterni, valutando gli esiti per migliorare continuamente il proprio comportamento.

Velocità, autonomia e over-permissioning

Le stesse caratteristiche che rendono gli agenti AI così interessanti per il business rappresentano però un serio

motivo di preoccupazione per i CISO.

Il limitato livello di supervisione applicato a questi sistemi crea maggiori opportunità di compromissione prima ancora che il team IT si accorga di un'anomalia. Inoltre, poiché gli agenti operano trasversalmente su più domini e applicazioni, aumenta il rischio di attacchi indiretti tramite prompt injection.

Per un threat actor può bastare individuare le fonti informative da cui l'agente estrae dati e inserire al loro interno prompt malevoli nascosti.

Il problema si amplifica ulteriormente perché gli agenti AI possono avere accesso a informazioni aziendali e personali altamente sensibili, oltre alla capacità di prendere decisioni con effetti irreversibili.

Anche senza un attacco esterno, esiste poi il rischio della cosiddetta "excessive agency": un'eccessiva autonomia che può portare a risultati indesiderati, come la cancellazione di dati critici o l'utilizzo incontrollato di API costose. Errori singoli possono propagarsi rapidamente tra agenti autonomi, generando effetti a catena e danni sistemici.

In questo scenario, l'identità digitale è il vero punto nevralgico del rischio.

Se un agente dispone di privilegi eccessivi, può inconsapevolmente fornire accesso a dati sensibili a cybercriminali o utenti interni a basso privilegio. Oppure può modificare, eliminare o utilizzare informazioni in modi non previsti.

La differenza rispetto alla tradizionale AI generativa è sostanziale: gli agenti AI sono persistenti, autonomi e operano alla velocità della macchina.

Shadow AI e rischi già concreti

A complicare ulteriormente il quadro c’è il fenomeno della Shadow AI. IBM stima che il 20% delle organizzazioni globali abbia subito nell’ultimo anno una violazione dei dati legata a strumenti AI non gestiti, con un incremento medio dei costi di breach pari a 675.000 dollari.

E i rischi non sono più teorici.

Uno degli esempi più evidenti è OpenClaw, assistente di AI agentica che ha rapidamente guadagnato popolarità alla fine del 2025. Delegando autorizzazioni su applicazioni SaaS, piattaforme cloud, sistemi di messaggistica e sessioni browser, OpenClaw rappresenta un potenziale tesoro per attacchi di credential theft, lateral movement e business email compromise.

A peggiorare il quadro, alcuni ricercatori hanno individuato vulnerabilità sia nella piattaforma stessa sia nei componenti aggiuntivi distribuiti tramite repository esterni.

Anche grandi aziende tecnologiche hanno sperimentato conseguenze inattese. Un malfunzionamento di OpenClaw ha quasi causato la cancellazione completa della casella email del responsabile AI safety di Meta. In un altro episodio, un agente AI ha diffuso informazioni interne riservate a personale non autorizzato dopo aver fornito indicazioni errate a un dipendente.

Persino AWS ha registrato problemi significativi: agenti incaricati di risolvere automaticamente anomalie operative hanno finito per causare interruzioni cloud protratte per ore.

Un ulteriore caso emblematico arriva da ServiceNow, dove alcuni ricercatori hanno scoperto che gli agenti AI utilizzavano un secret condiviso e non soggetto a rotazione. Questo ha consentito loro di ottenere privilegi amministrativi

all’interno di un’istanza della piattaforma.

È la dimostrazione di come pratiche IAM deboli possano diventare ancora più pericolose quando vengono integrate in ecosistemi AI.

Governance e identity security come priorità

Nel lungo periodo, standard e framework normativi contribuiranno probabilmente a regolamentare l’utilizzo dell’AI. Oggi, però, molte organizzazioni stanno adottando queste tecnologie sperando semplicemente che nulla vada storto.

È necessario fare di più. Il primo passo deve essere la governance. Come suggerisce Forrester, le aziende dovrebbero innanzitutto costruire inventari completi degli agenti AI in uso, comprendendo dove operano e quali funzioni svolgono. Solo dopo è possibile definire policy di utilizzo accettabile e valutare quali casi d’uso siano troppo rischiosi.

Dal punto di vista tecnologico, servono controlli endpoint evoluti come EDR/XDR, insieme a strumenti PAM e MDM capaci di limitare le azioni consentite agli utenti e agli agenti sui dispositivi aziendali.

Nel lungo termine, anche l’analisi comportamentale potrà giocare un ruolo fondamentale, aiutando i sistemi a distinguere se determinate attività siano eseguite da esseri umani o da agenti AI, e applicando restrizioni automatiche quando necessario.

Per gli ambienti Microsoft Entra ID, workflow di consenso e policy di conditional access rappresentano già oggi strumenti efficaci per limitare i rischi legati alle workload identity e ai service principal utilizzati dagli agenti.

Rallentare per proteggersi

Il problema principale resta però la velocità con cui le aziende vogliono adottare l’intelligenza artificiale.

Nella corsa al vantaggio competitivo, la sicurezza rischia di rimanere indietro. Per questo motivo, una delle azioni più importanti potrebbe essere semplicemente rallentare.



Consentire ai team di sicurezza di comprendere realmente quali tecnologie vengono introdotte, quali privilegi ricevono e quali rischi comportano è oggi fondamentale per prendere decisioni consapevoli e basate sul rischio.

Perché senza una solida strategia di identity security, gli agenti AI potrebbero trasformarsi rapidamente da acceleratori di business ad acceleratori di compromissione.

Articolo a cura di di Eric Woodruff, Chief Identity Architect, Semperis

Condividi sui Social Network:
https://www.ictsecuritymagazine.com/notizie/ai-agentici-identity-security/