Sicurezza IA: il tuo vantaggio competitivo nella corsa all'intelligenza artificiale

Una guida per CxO per accelerare l'adozione dell'IA gestendo al contempo i rischi

Introduzione

Mentre le organizzazioni si affrettano a sfruttare il potere trasformativo dell'IA, l'adozione spesso prevale sulla sicurezza. Per ogni iniziativa approvata, esistono innumerevoli casi di shadow AI creati da singoli e team. Infatti, l'85% dei responsabili delle decisioni IT afferma che i dipendenti stanno adottando strumenti di IA prima ancora che i loro team IT possano valutarli.[1] E il 93% dei dipendenti ammette di inserire informazioni negli strumenti IA senza approvazione.[1]

Nel frattempo, i vettori di attacco nativi dell'IA sono in forte aumento, con un incremento del 47% nell'ultimo anno[2], poiché gli strumenti di sicurezza tradizionali faticano a tenere il passo. Le nuove forme di esposizione dei dati e le lacune di conformità rappresentano una sfida per le prassi di governance esistenti. La dirigenza si trova di fronte a una domanda cruciale: come gestire i rischi introdotti dall'IA senza ostacolare l'innovazione che promette?

I requisiti sono chiari. Le organizzazioni devono creare un ambiente in cui:

- Tutto l'utilizzo dell'IA sia noto
- Tutte le comunicazioni dell'IA possano essere protette
- Tutti i criteri IA possano essere applicati
- Tutti i modelli IA possano essere protetti dagli abusi

Cloudflare AI Security Suite offre alle organizzazioni la sicurezza per accelerare l'adozione dell'IA, eliminando l'incertezza legata ai rischi. La nostra piattaforma unificata protegge l'intero ciclo di vita dell'IA, rilevando l'utilizzo dell'IA, applicando un accesso Zero Trust e proteggendo gli endpoint web e API con una difesa proattiva dalle minacce. Grazie alla governance dei dati integrata, i team possono innovare liberamente senza compromettere la sicurezza.

85%
dei responsabili delle decisioni IT afferma che i dipendenti stanno adottando strumenti IA prima ancora che i loro team IT possano valutarli.[1]

La sfida alla sicurezza dell'IA

I rischi dell'adozione dell'IA alimentano un mercato in rapida crescita per le soluzioni di sicurezza dell'IA. Ad esempio, le piattaforme di protezione delle applicazioni cloud native (CNAPP) si sono concentrate sui flussi di lavoro di sviluppo dell'IA. Le prime offerte come queste sono una linea di difesa essenziale, ma non sono sufficienti. Le aziende devono proteggere l'intero ciclo di vita dell'IA, inclusa la protezione dei sistemi IA una volta che sono in esecuzione in produzione.

La piena portata della sfida alla sicurezza dell'IA sta diventando fin troppo chiara. Oggi, gli sviluppatori creano funzionalità IA, i dipendenti utilizzano strumenti IA esterni e i clienti interagiscono con applicazioni basate sull'IA. Proteggere manualmente tutti questi ambienti eterogenei è complesso, e farlo in modo coerente è ancora più difficile.

Prevenzione dell'uso improprio dell'IA da parte della forza lavoro

Man mano che la forza lavoro aziendale adotta strumenti IA, sia autorizzati che non autorizzati, i dati e le operazioni sensibili sono a rischio. I team di sicurezza devono affrontare:

- **Mancanza di visibilità sull'utilizzo dell'IA:** spesso, i responsabili non hanno un quadro completo degli strumenti IA utilizzati dai dipendenti, dei luoghi in cui vengono elaborati i dati sensibili e delle modalità di connessione di questi sistemi IA alle applicazioni aziendali. Questo punto cieco crea una notevole esposizione al rischio.
- **Rischi per la sicurezza e la conformità dei dati:** l'IA cambia il modo in cui i dati transitano nella tua organizzazione. Le informazioni personali, i dati proprietari e le registrazioni dei clienti possono finire nei sistemi IA in modi che causano violazioni della conformità o espongono informazioni riservate.
- **Controlli di accesso per i flussi di lavoro dell'IA agentica:** non è solo l'accesso umano agli strumenti IA che deve essere gestito. È necessario gestire anche l'accesso dell'IA agentica ai server MCP e ad altri sistemi critici. Questo richiede nuovi approcci alla gestione delle identità e ai controlli di accesso.

Protezione di applicazioni e modelli IA rivolti al pubblico

Che siano sviluppati internamente o consumati tramite terzi, i sistemi IA stanno diventando un pilastro essenziale dell'esperienza del cliente e dell'utente e, come tali, devono essere protetti. La Top 10 di OWASP per le applicazioni LLM mette in evidenza le minacce che gli strumenti tradizionali non sono in grado di affrontare:

- **Attacchi di consumo illimitato:** simili ai tradizionali attacchi DoS, gli attacchi di consumo illimitato cercano di sopraffare gli LLM con richieste a uso intensivo di risorse. I modelli di prezzo cloud pay-per-use possono far aumentare notevolmente l'impatto finanziario di tali attacchi, mentre gli utenti legittimi subiscono un peggioramento della qualità del servizio.
- **Model poisoning:** gli autori degli attacchi impiantano backdoor, bias o vulnerabilità negli LLM inoculando dati danneggiati in set di dati pubblici o repository utilizzati dagli sviluppatori per l'addestramento dei modelli. Un modello "avvelenato" in questo modo si comporta normalmente fino a quando specifici trigger non attivano comportamenti dannosi.
- **Attacchi di inoculazione di prompt:** una tattica popolare per l'esfiltrazione dei dati tramite interfacce IA. L'inoculazione di prompt manipola gli input degli LLM incorporando istruzioni dannose all'interno dei prompt dell'utente o di contenuti esterni, inducendo i modelli a ignorare le istruzioni originali e a eseguire invece i comandi dell'aggressore.
- **Jailbreaking:** prompt accuratamente elaborati, come scenari di role-playing, comandi di override delle istruzioni e strategie multi-turn, possono aggirare le protezioni di sicurezza degli LLM per generare contenuti vietati o estrarre informazioni sensibili.

Per consentire una rapida innovazione dell'IA senza mettere a rischio l'organizzazione, si consiglia ai responsabili di adottare un approccio alla sicurezza dell'IA che comprenda l'intero spettro dei requisiti:

- Protezione dell'utilizzo della GenAI da parte della forza lavoro
- Protezione di app e carichi di lavoro basati sull'IA
- Creazione di app basate sull'IA sicure fin dalla progettazione

Protezione degli ambienti di sviluppo e dei flussi di lavoro di addestramento dell'IA

I progetti IA implicano set di dati massicci, risorse costose e sperimentazioni iterative, creando nuove superfici d'attacco e vulnerabilità. I team di sicurezza devono affrontare:

- **Sicurezza e integrità dei dati di addestramento:** dati di addestramento compromessi possono introdurre bias, backdoor o vulnerabilità che persistono nei modelli di produzione. Le organizzazioni devono proteggere l'accesso ai set di dati di addestramento, prevenire modifiche non autorizzate e garantire il throughout della provenienza dei dati durante l'intero ciclo di vita del modello.
- **Gestione di credenziali e segreti:** i flussi di lavoro di sviluppo dell'IA richiedono l'accesso a sistemi quali archiviazione cloud per set di dati, cluster di elaborazione per l'addestramento, registri di modelli e servizi di terzi. Segreti o chiavi API non adeguatamente protette possono esporre credenziali sensibili, consentendo l'accesso non autorizzato a modelli proprietari, dati di addestramento o sistemi di produzione.
- **Controlli di accesso all'ambiente di sviluppo:** i tecnici IA spesso necessitano di privilegi elevati per sperimentare con i modelli e accedere a dati sensibili. Senza adeguati controlli degli accessi, questo può causare minacce interne, esposizione accidentale dei dati o estrazione non autorizzata di modelli.

Unificare la sicurezza con Cloudflare

L'adozione efficace dell'IA si concentra sull'abilitazione della produttività, non sulla sua limitazione. Quando i team possono utilizzare l'IA con sicurezza, sapendo che sono state implementate le misure di sicurezza adeguate, innovano più rapidamente e intraprendono progetti più ambiziosi.

Cloudflare AI Security Suite crea un ambiente sicuro per l'innovazione dell'IA. Unificando le funzionalità di Secure Access Service Edge (SASE) e di sicurezza delle applicazioni Web, i CxO possono connettere e proteggere due domini fondamentali:

- Applicazioni esterne rivolte al pubblico basate sull'IA
- Sistemi e carichi di lavoro IA interni e privati

Concentrandosi sull'intero ciclo di vita dell'IA, Cloudflare AI Security Suite soddisfa le esigenze di sicurezza, dall'individuazione e gestione dei rischi alla protezione dei dati, garantendo l'accesso sicuro agli utenti e proteggendo le applicazioni basate sull'IA e i flussi di lavoro di sviluppo. Per fornire un livello essenziale di sicurezza della produzione in tempo reale, la rete globale di Cloudflare si interpone per ispezionare e filtrare ogni interazione dell'IA, proteggendo i dati di tutti gli utenti e le applicazioni.

Invece di limitarsi a rilevare i problemi quando si verificano, Cloudflare previene i problemi e blocca le minacce prima che raggiungano i modelli IA. I team di sicurezza ottengono la visibilità necessaria per stare al passo con le minacce emergenti.

Funzionalità principali di Cloudflare

Cloudflare AI Security Suite offre un monitoraggio completo, protezione in tempo reale e gestione proattiva del rischio, tutto sotto un unico ombrello, per un approccio olistico alla sicurezza dell'IA.

Rilevamento e visibilità completi dell'IA

Una sicurezza efficace dell'IA dipende da un inventario completo e in tempo reale delle risorse e dell'utilizzo dell'IA, sia autorizzati che non autorizzati. Il fondamento di Cloudflare AI Security Suite è il monitoraggio continuo e il rilevamento automatizzato per identificare tutti i modelli IA, gli assistenti, gli agenti e le distribuzioni di shadow AI in ogni tipo di ambiente: pubblico, privato o interno.

Gestione proattiva del rischio dell'IA

Cloudflare AI Security Suite aiuta le organizzazioni a prevenire gli attacchi rilevando e mitigando vulnerabilità, errori di configurazione e percorsi di attacco specifici per l'IA, compresi quelli presenti nella Top 10 OWASP per gli LLM. Il punteggio di affidabilità delle applicazioni aiuta a stabilire le priorità di correzione, in modo che i team affrontino innanzitutto i rischi più significativi.

Sicurezza delle applicazioni per app basate sull'IA

Per aiutare le SecOps a rimanere aggiornate sui più recenti vettori di minaccia dell'IA, Cloudflare AI Security Suite include il rilevamento e la mitigazione proattivi delle minacce per vulnerabilità specifiche dell'IA, errori di configurazione e percorsi di attacco all'interno delle pipeline IA, compresa la protezione da attacchi di inoculazione di prompt, data poisoning e abuso di modelli.

I firewall IA specializzati rilevano ed etichettano gli endpoint API e di IA generativa o agentica, rilevano i tentativi di esfiltrazione di informazioni di identificazione personale (PII) e bloccano i prompt dannosi prima che influiscano sulle prestazioni dei modelli IA o causino model poisoning con contenuti tossici o disinformazione.

Accesso Zero Trust per i flussi di lavoro di GenAI e IA agentica

I principi Zero Trust, come il principio del privilegio minimo, si applicano sia alla forza lavoro che agli agenti IA. Cloudflare AI Security Suite è in grado di applicare criteri Zero Trust Network Access (ZTNA) sia per le interazioni uomo-IA che IA-IA. La registrazione e i controlli centralizzati per i server MCP assicurano che l'IA agentica acceda solo a ciò per cui è stata autorizzata, inclusi i flussi di lavoro just-in-time.

Protezione dei dati basata sull'IA

Una sicurezza dell'IA efficace richiede una funzionalità di prevenzione della perdita di dati in grado di sfruttare più modelli linguistici per comprendere sia il contenuto di un prompt sia l'obiettivo alla sua base. Cloudflare AI Security Suite integra funzionalità di prevenzione della perdita di dati (DLP) durante l'addestramento, i prompt e le risposte, al fine di prevenire l'esposizione di informazioni di identificazione personale (PII), la fuga di dati e l'accesso non autorizzato all'interno di modelli e pipeline IA. La sicurezza di runtime distribuita in linea e incentrata sulle API funge da primo livello di difesa rapido e semplice, a complemento dell'approccio shift-left abilitato da un CNAPP.

Localizzazione dei dati

Gli LLM e le applicazioni IA sono soggette alle stesse normative di qualsiasi altro ambiente di dati. Cloudflare AI Security Suite aiuta le organizzazioni a garantire che i carichi di lavoro IA rispettino i confini geografici e giurisdizionali tramite criteri che mantengono i dati di addestramento e le richieste di inferenza all'interno delle aree geografiche approvate.

Sicurezza fin dalla progettazione per lo sviluppo dell'IA

Come per tutti i tipi di software, la sicurezza delle app IA deve essere integrata fin dall'inizio del ciclo di sviluppo, e non aggiunta in un secondo momento. Cloudflare AI Security Suite fornisce agli sviluppatori gli strumenti e i framework necessari per creare app basate sull'IA sicure fin

L'impatto aziendale di Cloudflare AI Security Suite

L'ascesa dell'IA non è solo un'evoluzione, è una svolta epocale paragonabile all'industrializzazione o all'informatizzazione. In tal modo, l'adozione rapida ed efficace diventa cruciale non solo per stimolare la crescita, ma anche per la sostenibilità organizzativa. L'adozione lenta o non sicura rappresenta ormai una minaccia esistenziale per le organizzazioni in tutti i settori e comparti.

Cloudflare AI Security Suite consente una trasformazione sicura, controllata ed efficiente per soddisfare le crescenti aspettative dei clienti e le esigenze del mercato in ambienti estremamente competitivi.

- **Innovazione dell'IA più rapida:** quando la sicurezza consente un utilizzo sicuro anziché bloccare l'adozione, i dipendenti e i team possono esplorare nuove app IA per diventare più produttivi nel loro lavoro quotidiano. Framework di sicurezza adeguati offrono agli sviluppatori la sicurezza di creare funzionalità IA ambiziose senza compromettere dati o sistemi sensibili.
- **Rischio correlato all'IA ridotto:** le organizzazioni possono gestire l'intero spettro di rischi derivanti dall'adozione dell'IA, compresi i rischi legati all'IA inerenti alle applicazioni Web abilitate per l'IA. Le app IA in fase di sviluppo possono essere protette per garantire che i dipendenti non divulghino dati sensibili o li espongano nei set di addestramento dell'IA. Le SecOps possono identificare e mitigare le minacce e le vulnerabilità specifiche dell'IA in modo proattivo, riducendo al minimo la superficie d'attacco e proteggendo dati e modelli critici.
- **Operazioni di sicurezza semplificate:** la visibilità e il controllo centralizzati sullo stato della sicurezza dell'IA semplificano la gestione e ottimizzano la risposta agli incidenti. Il tuo team SecOps può concentrarsi su iniziative strategiche, anziché gestire costantemente gli incidenti relativi all'IA.
- **Governance e conformità dei dati solide:** i controlli di protezione dei dati specifici per l'IA ti aiutano a proteggere le informazioni sensibili e a soddisfare i requisiti normativi in continua evoluzione durante l'intero ciclo di vita dell'IA.
- **Costo totale di proprietà (TCO) inferiore:** sfruttare una piattaforma consolidata che integra gli investimenti di sicurezza esistenti è più economicamente vantaggioso rispetto all'implementazione di soluzioni monofunzionali separate per ogni problematica di sicurezza dell'IA.

Casi d'uso del modello per Cloudflare AI Security Suite

Cloudflare AI Security Suite è ideale per soddisfare i requisiti fondamentali relativi all'adozione dell'IA.

- **Protezione dell'utilizzo degli strumenti IA della forza lavoro:** applica criteri Zero Trust per l'accesso dei dipendenti sia a strumenti di IA generativa pubblici come ChatGPT, sia ad app basate sull'IA sviluppate internamente.
- **Protezione di app rivolte al pubblico e abilitate per l'IA:** proteggi applicazioni web e API in cui sono integrati modelli IA, come chatbot e motori di suggerimenti, da attacchi che potrebbero esporre dati sensibili o compromettere il modello.
- **Gestione della shadow AI:** scopri automaticamente gli strumenti IA non autorizzati nella tua organizzazione e applica i controlli appropriati per consentire una continua innovazione entro i confini di rischio gestiti.
- **DLP basata sull'IA per le interazioni con l'IA:** impedisce che i dati sensibili vengano esposti nei prompt o nelle risposte dell'IA, garantendo che le informazioni di identificazione personale (PII) e le informazioni riservate rimangano protette.
- **Sicurezza dello sviluppo dell'IA:** fornisci ai team di tecnici framework che rendano lo sviluppo sicuro l'approccio predefinito, consentendo loro di creare rapidamente funzionalità IA senza compromettere la sicurezza.

Considerazioni per l'implementazione

Come strategia di sicurezza fondamentale, la sicurezza dell'IA deve essere affrontata con attenzione.

Inizia con la tua infrastruttura esistente	Le implementazioni di sicurezza dell'IA di maggior successo si basano sugli strumenti SASE e di sicurezza delle applicazioni esistenti, anziché sostituirli. Questo approccio sfrutta gli investimenti attuali, estendendo al contempo la protezione per coprire i rischi specifici dell'IA.
Distribuisci protezioni in linea unificate	La capacità di bloccare le attività dannose al perimetro di rete, in tempo reale, è cruciale per la sicurezza dell'IA. Distribuisci controlli in linea in tempo reale e completali con il monitoraggio basato su API.
Assicura una copertura completa	La tua strategia di sicurezza dell'IA deve affrontare l'intero spettro di requisiti: utilizzo da parte della forza lavoro di strumenti GenAI, protezione per app e flussi di lavoro basati sull'IA, protezione dei flussi di lavoro di IA agentica e sicurezza per i flussi di lavoro di sviluppo dell'IA.
Pianificazione su scala aziendale	Scegli soluzioni scalabili in base alla tua adozione dell'IA. Ciò che funziona per un programma pilota deve scalare anche a livello dell'intera organizzazione con l'espandersi dell'utilizzo dell'IA.
Convalida i tuoi criteri di successo	Prima di impegnarsi completamente, è necessario convalidare la soluzione in scenari reali. Scegli una piattaforma che consenta l'attivazione gratuita e self-service delle sue funzionalità aziendali. Questo ti consente di testare l'intera suite di sicurezza su piccola scala, ad esempio per un singolo team o app, per dimostrarne rapidamente il valore e assicurarti che soddisfi i tuoi criteri di successo.

Passare alla fase successiva con Cloudflare AI Security Suite

Con l'accelerazione dell'adozione dell'IA a livello globale, le organizzazioni in grado di scalare l'IA in modo sicuro otterranno un vantaggio competitivo significativo. La chiave è riconoscere che la sicurezza dell'IA non consiste nel prevenire l'uso dell'IA, bensì nell'abilitarla in modo intelligente.

Cloudflare AI Security Suite consente alle organizzazioni di innovare in tutta sicurezza. Basandosi ed estendendo le nostre piattaforme SASE e Application Security ampiamente adottate, la nostra AI Security Suite offre rilevamento integrato dell'IA, controlli di accesso Zero Trust, difesa proattiva dalle minacce basata sull'intelligence e una solida governance dei dati. Proteggendo l'utilizzo e i modelli dell'IA da ogni punto di vista, le organizzazioni possono consentire agli sviluppatori di creare in tempi più rapidi e rendere i dipendenti più produttivi, il tutto senza compromettere l'esperienza dell'utente finale.

Prenota una consulenza

Scopri come Cloudflare AI Security Suite può trasformare l'approccio della tua organizzazione all'adozione sicura dell'IA.

+44 (0) 20 3514 6970

enterprise@cloudflare.com

www.cloudflare.com

1. ManageEngine. The Shadow AI Surge in Enterprises: Insights from the US & Canada
2. Check Point Research. Q1 2025 Global Cyber Attack Report from Check Point Software: An Almost 50% Surge in Cyber Threats Worldwide, with a Rise of 126% in Ransomware Attacks