La falla nella sicurezza dell'AI: una guida all'implementazione dello zero trust per i team di sicurezza

Una guida pratica per proteggere l'intero ciclo di vita dell'AI, dal rilevamento della Shadow AI alla protezione del runtime degli agenti con i principi zero trust, un piano d'implementazione a 30/60/90 giorni e indicatori KPI per i report destinati alla dirigenza.

E-BOOK

Introduzione

I team di sicurezza sanno di avere un problema con l'intelligenza artificiale. Quello che molti ancora non riescono a quantificare è la sua portata.

Quali applicazioni di intelligenza artificiale (AI) sono attualmente in esecuzione nel tuo ambiente? Quali modelli vengono interrogati dai tuoi sviluppatori tramite plugin per gli ambienti di sviluppo integrato (IDE) e gli strumenti della riga di comando (CLI)? Quali dati vengono trasferiti tramite i prompt? E il loro contenuto include codice sorgente, informazioni personali dei clienti o contenuti soggetti a obblighi normativi? Quali agenti agiscono autonomamente per conto degli utenti e a quali sistemi possono accedere?

È proprio nel divario tra ciò che l'AI fa nel tuo ambiente e ciò che sei in grado di monitorare e governare che si annida il rischio reale. Non in scenari di attacco teorici, ma nell'uso quotidiano, non autorizzato e in buona fede di strumenti AI che la sicurezza non ha mai approvato perché nessuno lo ha richiesto. La Shadow AI e le funzionalità AI integrate in piattaforme SaaS (Software as a Service) già approvate sono le principali fonti di questo divario. Uno strumento approvato dal reparto IT per uno scopo specifico che introduce un assistente AI sempre attivo che elabora i dati interni senza alcuna verifica di sicurezza.

I dati e le statistiche presenti in questa guida sono tratti dal Report sulla sicurezza dell'AI del 2026 di ThreatLabz, pubblicato da Zscaler nel 2026 e basato sull'analisi di 989,3 miliardi transazioni AI/ML osservate sulla piattaforma Zscaler Zero Trust Exchange da gennaio a dicembre 2025. Tutti i dati si riferiscono al periodo di ricerca, salvo diversa indicazione.

Il modello zero trust offre una risposta basata su principi solidi:

Non considerare alcuna app, utente, richiesta o connettore intrinsecamente affidabile. Applica l'accesso basato sull'identità e sul contesto a ogni livello. Verifica tutto continuamente. Riduci al minimo il raggio d'azione delle potenziali violazioni. La difficoltà sta nello stabilire l'ordine delle priorità. La visibilità deve venire prima dell'applicazione delle policy. I controlli di accesso devono essere stabili prima di implementare l'automazione. Se si procede in modo non sequenziale, ci si ritrova con controlli che bloccano il lavoro legittimo o non riescono a prevenire i rischi reali.

83%	**18.033 TB**	**39%**	**93%**
aumento annuo delle transazioni AI/ML nelle aziende: l'adozione è ormai su scala infrastrutturale	di dati trasferiti tramiteapp di intelligenza artificiale nel 2025: più di quanto la maggior parte dei programmi sia in grado di tracciare	di transazioni AI attivamente bloccate: la pressione per una governance efficace è già presente	Aumento annuo del volume dei dati trasferiti : superiore alla capacità dei programmi di sicurezza

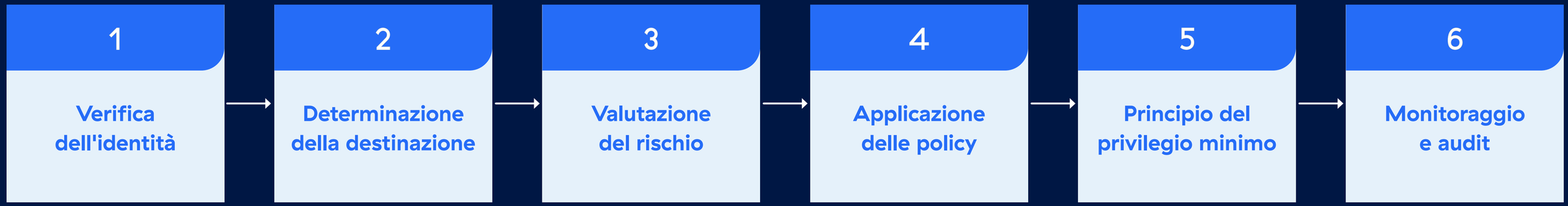
Fonte: Report sulla sicurezza dell'AI del 2026 di ThreatLabz, Zscaler (pp. 5—6)

Cosa tratta questa guida

1. I sette principi dello zero trust applicati nello specifico a sistemi, agenti e flussi di dati AI
2. Un piano d'implementazione graduale a 30/60/90 giorni definito in base alle dipendenze
3. Copertura degli ambiti di rischio emergenti legati all'AI: Shadow AI, attacchi di prompt injection, sistemi agentici, traffico Model Context Protocol (MCP)/Agent-to-agent (A2A), AI incorporata in strumenti SaaS e profilo di sicurezza dell'infrastruttura AI.
4. Un modello KPI per presentare alla dirigenza il progresso del programma di sicurezza AI

Principi zero trust applicati all'intelligenza artificiale

I sette principi delineati in seguito costituiscono il quadro decisionale alla base di tutte le attività previste nel piano d'azione dell'implementazione. Ciascuno di essi è uno strumento decisionale per la valutazione dei controlli. La sequenza rispecchia il modo in cui un team di sicurezza li applicherebbe effettivamente nella pratica.



Applicazione in sequenza. Ogni principio si basa su quello precedente.

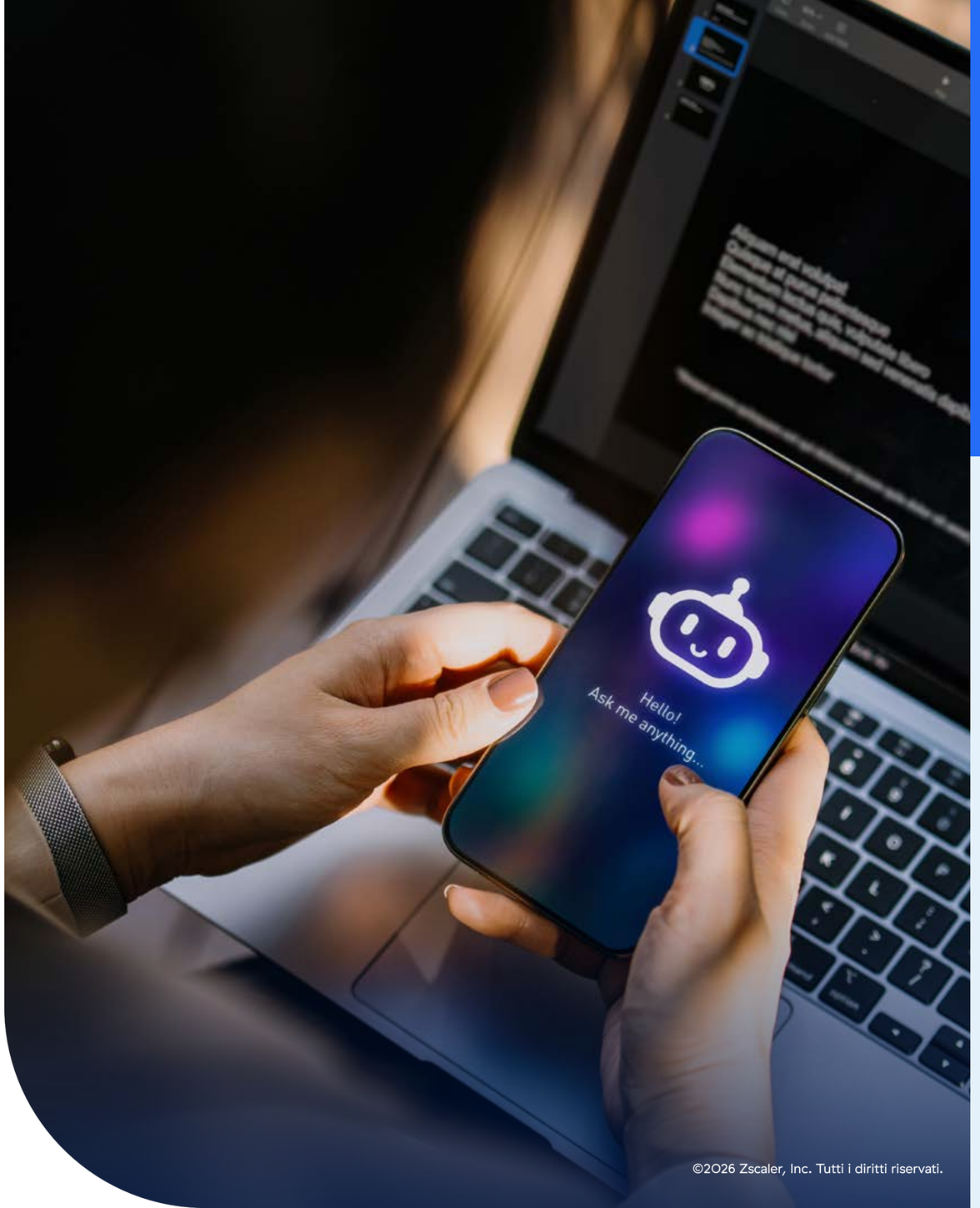
Principio 1:
VERIFICA L'IDENTITÀ DI UTENTI, SVILUPPATORI E SERVIZI

L'accesso all'intelligenza artificiale inizia con l'identità. Ogni utente che accede a uno strumento SaaS di intelligenza artificiale generativa (GenAI), ogni sviluppatore che interroga un endpoint del modello da un plugin IDE e ogni account di servizio che si connette a un componente dell'infrastruttura AI dovrebbe autenticarsi tramite un piano di controllo dell'identità coerente. Il Single Sign-On (SSO) e l'autenticazione a più fattori (MFA) sono standard di base, ma i dettagli sono importanti.

I dipendenti e gli sviluppatori presentano profili di rischio differenti e richiedono un trattamento distinto nelle policy. Un analista di marketing che utilizza ChatGPT per creare bozze di contenuti presenta rischi di esposizione dei dati diversi rispetto a un ingegnere software il cui IDE instrada i completamenti del codice tramite un plugin GitHub Copilot connesso a un server MCP con accesso a repository interni. L'autenticazione tra servizi per gli agenti AI è il caso limite emergente per il quale la maggior parte delle organizzazioni non è ancora preparata.

Controlli principali

- Applicazione di SSO e MFA per le app SaaS e gli strumenti di sviluppo AI
- Policy di identità separate per l'utilizzo della GenAI da parte del personale rispetto agli strumenti AI per sviluppatori (IDE, CLI, integrazioni MCP).
- **Controlli degli account di servizio per gli agenti AI:** privilegi minimi, credenziali di breve durata e registrazione degli eventi per gli audit.
- **Il profilo di sicurezza del dispositivo come segnale di identità:** gli endpoint gestiti ottengono un accesso diverso rispetto ai dispositivi personali (BYOD).



Principio 2:
DETERMINARE LA DESTINAZIONE PER L'AI AUTORIZZATA E NON AUTORIZZATA

Non tutte le applicazioni di intelligenza artificiale sono uguali, e un approccio indiscriminato che le blocchi tutte non è né realistico dal punto di vista operativo né efficace dal punto di vista della sicurezza. **L'obiettivo è una policy di destinazione a più livelli:** gli strumenti autorizzati ottengono l'accesso totale con ispezione inline, gli strumenti tollerati ottengono un accesso monitorato con limiti di utilizzo, mentre gli strumenti non autorizzati o ad alto rischio vengono bloccati o isolati. L'elenco delle superfici che interagiscono con l'AI è più ampio di quanto la maggior parte delle organizzazioni si aspetti: in un solo anno, infatti, con l'integrazione di nuove funzionalità AI nelle piattaforme esistenti, si è quadruplicato.

Controlli principali

- Policy di destinazione a più livelli, per consentire l'uso, inviare avvisi ed eseguire azioni di isolamento o blocco in base all'app o al gruppo di utenti
- Rilevamento dinamico delle app per individuare i nuovi strumenti AI non appena compaiono nel traffico
- Percorsi di policy separati per le app SaaS di AI generativa, l'AI incorporata nelle app SaaS usate dall'azienda e gli strumenti di sviluppo AI
- Flusso di approvazione per le richieste di nuovi strumenti AI, in modo che gli utenti seguano un percorso predefinito e non aggirino i controlli

OLTRE 3.400

APPLICAZIONI AI DIVERSE TRACCIATE NEL TRAFFICO DELLE AZIENDE: UN AUMENTO DI 4 VOLTE IN UN SOLO ANNO E CONTINUAMENTE IN CRESCITA

Report sulla sicurezza dell'AI di ThreatLabz del 2026, Zscaler (pag. 5)

Principio 3:
VALUTARE IL RISCHIO CON DECISIONI BASATE SUL CONTESTO

Le decisioni di accesso all'AI non dovrebbero mai ridursi a un semplice sì o no. Lo stesso utente che utilizza lo stesso strumento di intelligenza artificiale può comportare rischi diversi in base al contesto operativo, ovvero il dispositivo, la sensibilità dei dati e l'azione intrapresa. Per effettuare una valutazione efficace del rischio, è fondamentale identificare in anticipo i dati più sensibili, come il codice sorgente, i dati personali, i dati relativi alle carte di pagamento, le informazioni sanitarie, i risultati finanziari non ancora pubblicati e le informazioni riservate relative a fusioni e acquisizioni. Le principali categorie di violazioni rilevate nel traffico aziendale mostrano esattamente quali dati i dipendenti inviano: nomi e dati identificativi nazionali, codice sorgente e cartelle cliniche.

Controlli principali

- Categorie esplicite di dati critici per il codice sorgente, i dati personali, sanitari e delle carte di pagamento, i segreti aziendali e i piani ancora non divulgati
- Valutazione del rischio per gruppo di utenti, reparto, profilo del dispositivo e sensibilità dei dati
- Adeguamento dinamico delle policy in base a segnali contestuali, non solo a regole statiche di autorizzazione/blocco
- Gestione separata del rischio per le interazioni con l'AI che includono l'upload di file rispetto a quelle basate esclusivamente su prompt testuali

OLTRE 410 MILIONI

DI VIOLAZIONI DELLA DLP LEGATE SOLO A CHATGPT: FUGA DI NOMI, NUMERI DI PREVIDENZA SOCIALE, CODICE SORGENTE, DATI MEDICI.

Report sulla sicurezza dell'AI di ThreatLabz del 2026, Zscaler (pp. 6, 15).

Principio 4:

APPLICAZIONE DELLE POLICY INLINE PER OGNI SESSIONE

I log ti dicono cosa è successo. Quando trovi un record che indica che uno sviluppatore ha inviato codice sorgente proprietario a un modello AI non autorizzato, i dati sono già stati compromessi. La maggior parte delle organizzazioni peggiora ulteriormente questo problema implementando soluzioni DLP per l'upload dei file verso gli strumenti AI, lasciando però i prompt di testo privi di controlli. Uno sviluppatore che digita il codice sorgente direttamente in una finestra di prompt aggira questi controlli senza che venga generato alcun avviso. L'aumento dei dati trasferiti mostra la portata di ciò che transita attraverso questi canali non protetti.

Controlli principali

- DLP inline su prompt, risposte e upload di file, non solo a livello di URL/categoria
- Azioni di autorizzazione, avviso, oscuramento e isolamento disponibili nei punti di applicazione delle policy
- Classificazione rapida per le principali app di GenAI con rilevamento del contesto
- Moderazione dei contenuti per interazioni off topic, soggette a restrizioni e sensibili per la concorrenza
- Isolamento del browser per gli strumenti AI consentiti ma rischiosi: accesso senza piena fiducia

93%

AUMENTO ANNUO DEL VOLUME DI DATI AZIENDALI TRASFERITI TRAMITE L'AI, SUPERIORE A QUANTO LA MAGGIOR PARTE DEI PROGRAMMI DI SICUREZZA È IN GRADO DI TRACCIARE O GESTIRE.

Report sulla sicurezza dell'AI di ThreatLabz del 2026, Zscaler (pp. 5, 14)

Principio 5:
MINIMIZZAZIONE DI ACCESSO ED ESPOSIZIONE

I sistemi AI creano percorsi di accesso che spesso si collocano al di fuori dei modelli tradizionali di identità e rete. Una pipeline RAG (Retrieval–Augmented Generation) può consentire a un modello di accedere a un insieme di documenti. Un agente può connettersi a strumenti interni, dataset, archivi vettoriali o server MCP. Queste autorizzazioni sono effettive quanto quelle di un utente reale, anche se non risultano nelle revisioni degli accessi standard.

Ridurre al minimo l'accesso e l'esposizione significa limitare ciò che i sistemi di intelligenza artificiale possono raggiungere e restringere le modalità di comunicazione tra i componenti AI. Agenti, pipeline di dati, componenti RAG e fonti dati di backend devono connettersi solo tramite percorsi esplicitamente approvati. Senza un'analisi e una segmentazione basate sul principio del privilegio minimo, un agente con autorizzazioni eccessive o un componente AI compromesso possono diventare un percorso di movimento laterale.

Controlli principali

- Definizione degli ambiti di accesso in base ai ruoli per app AI, agenti, connettori di modelli, dataset, archivi vettoriali e componenti dell'infrastruttura AI
- Revisione periodica e riduzione dell'accesso non autorizzato per dataset di informazioni sensibili e sistemi connessi all'intelligenza artificiale
- Autorizzazioni minime per gli agenti AI: definisci a cosa possono accedere, limita tutto il resto
- Segmentazione dei servizi AI, delle piattaforme di agenti, delle pipeline di dati, dei componenti RAG e delle fonti di dati di backend
- Limita l'accesso est–ovest tra i componenti AI a percorsi definiti esplicitamente
- Ispezione e registrazione basate su protocollo per il traffico MCP e A2A

Riduci al minimo il potenziale impatto lungo ogni percorso di accesso all'AI

Tratta ogni connettore AI, autorizzazione di agente, pipeline RAG, archivio vettoriale, server MCP e connessione tra componenti come un punto di esposizione. Definisci l'ambito di accesso, segmenta le aree di comunicazione, ispeziona il traffico e registra l'attività.

Principio 6:
MONITORAGGIO CONTINUO E VERIFICABILITÀ

La verifica continua è il modello operativo, non un traguardo da raggiungere e poi superare. Per i sistemi di intelligenza artificiale, questo si traduce nel logging di prompt e risposte, avvisi in tempo reale riguardo alla violazione delle policy, dashboard su cui visualizzare l'utilizzo e un flusso di lavoro di indagine che non richieda la correlazione manuale dei log su più strumenti diversi. **La pressione normativa in tal senso è reale:** la Legge sull'intelligenza artificiale dell'Unione Europea (EU AI Act), il Framework per la gestione del rischio dell'intelligenza artificiale del National Institute of Standards and Technology (NIST AI RMF) e i framework emergenti di governance dell'AI richiedono tutti controlli dimostrabili e documentati.

Controlli principali

- Logging di prompt e risposte per le interazioni AI inline
- Dashboard di utilizzo per utente, reparto e app AI con segnali di rischio integrati
- Flusso di lavoro di avviso e indagine per violazioni e anomalie delle policy relative all'AI
- Report pronti per gli audit: quali tecnologie AI vengono utilizzate, quali controlli vengono applicati e quali violazioni si sono verificate
- Mappatura della governance rispetto al NIST AI RMF, alla Legge sull'intelligenza artificiale dell'Unione Europea e ad altri framework applicabili

1	2	3	4	5
Reattivo	**Visibile**	**Governato**	**Automatizzato**	**Continuo**
Nessun inventario, Blocco ad hoc	Impronta AI conosciuta, Policy di base	Policy basate sui ruoli, Flussi di lavoro di governance	Coaching automatizzato, Conformità continua	Red teaming + protezione del runtime

Livello di maturità del programma →

Giorni 1–30

Stabilisci la visibilità

Non si può proteggere un'intelligenza artificiale che non si vede. Per i primi 30 giorni, concentrari sulla creazione di un inventario sufficiente e su un controllo di base per prendere decisioni informate, non sul blocco completo

GIORNI 1–30 Stabilisci la visibilità	GIORNI 31–60 Sviluppa la maturità delle policy	GIORNI 61–90 Automatizza e convalida
• Implementa il rilevamento delle app AI • Definisci le categorie di dati sensibili • Stabilisci le policy di accesso di base • Prima dashboard per la dirigenza	• Policy di accesso basate sui ruoli • Flussi di lavoro di governance in tempo reale • Implementazione di AI-SPM • Protezione dell'AI privata	• Coaching automatizzato degli utenti • Conformità continua • Cadenza delle attività di red teaming • Controlli di sicurezza a ciclo chiuso

Panoramica del piano di implementazione a 30/60/90 giorni

Checklist per l'implementazione in 30 giorni

- **Stabilisci una cadenza periodica per la revisione degli strumenti AI:** nuovi strumenti rilevati nel traffico, richieste di eccezione ricevute e modifiche apportate alle policy
- Crea la prima dashboard di sicurezza AI destinata alla dirigenza: impronta delle app AI (autorizzate vs non autorizzate), eventi DLP per categoria, azioni sulle policy e tendenze
- Definisci piani dettagliati di risposta agli incidenti per gli eventi di sicurezza AI più comuni: dati sensibili inseriti nei prompt, accesso a destinazioni bloccate e upload di file di grandi dimensioni negli strumenti AI
- **Centralizza i log di sicurezza dell'AI:** dati su attività degli utenti, utilizzo delle app, eventi di prompt/risposta, azioni sulle policy e violazioni disponibili in un'unica dashboard

Operatività

- **Rivedi e ottimizza le policy DLP utilizzando i dati di telemetria delle prime due settimane:** riduci i falsi positivi e migliora il rilevamento di modelli di esposizione dei dati confermati
- **Abilita la moderazione dei contenuti per garantire il rispetto delle regole di utilizzo accettabile:** interazioni off topic, argomenti soggetti a restrizioni, contenuti sensibili dal punto di vista della concorrenza e output inappropriati
- **Aggiungi l'isolamento del browser per gli strumenti AI nel livello consentito ma rischioso:** gli utenti possono accedere allo strumento, ma le attività di upload, download e copia/incolla sono limitati a livello del browser
- **Estendi la copertura DLP all'intero ambito dei dati sensibili:** codice sorgente (linguaggi e modelli aggiuntivi), dati personali (oltre i formati ovvi), dati dei titolari di carte PCI, informazioni sanitarie protette e contenuti aziendali riservati

Estendi la protezione dei dati

- **Pubblica una guida con un modello a tre colori sull'uso accettabile dell'AI:** verde (autorizzato), giallo (uso con restrizioni), rosso (bloccato, con motivazione). Rendi visibile il processo di richiesta di eccezione
- Abilita la protezione dei dati inline per i prompt a partire da regole di rilevamento ad alta affidabilità per le categorie di dati sensibili: modelli di codice sorgente, formati delle credenziali, schemi di informazioni personali
- **Limita le azioni ad alto rischio sugli strumenti non autorizzati:** upload, blocchi di copia/incolla di grandi dimensioni e download degli output dell'AI verso archivi non gestiti
- **Applica policy di destinazione:** avvisa o blocca in base al gruppo di utenti e all'app per destinazioni AI chiaramente non autorizzate o ad alto rischio

Applica i controlli di base

- **Stabilisci un inventario di base dei rischi legati all'AI:** quali app accedono a quali tipi di dati e quali gruppi di utenti utilizzano maggiormente l'intelligenza artificiale
- **Definisci esplicitamente le categorie di dati sensibili:** codice sorgente, dati personali, delle carte di pagamento e sanitari, segreti e credenziali, dati finanziari non pubblicati, informazioni riservate su fusioni e acquisizioni o piani strategici

Definisci cosa ha bisogno di protezione

- Abilita la visibilità immediata per le app AI con il traffico più elevato per comprendere quali dati transitano attraverso i prompt e le risposte
- Mappa gli strumenti AI che si integrano con i sistemi interni e a quali dati hanno accesso
- **Individua la Shadow AI:** strumenti utilizzati attivamente che non sono mai stati formalmente approvati.
- Implementa il rilevamento delle app AI per eseguire l'inventario della GenAI SaaS, dell'AI integrata in piattaforme aziendali e degli strumenti AI per sviluppatori in uso nell'ambiente

COMPROMESSI DA CONSIDERARE

Partire da policy di avviso anziché da blocchi definitivi riduce gli attriti e individua i falsi positivi nelle regole DLP prima che inizino a bloccare attività legittime. Lo svantaggio è che, durante la fase di avviso, possono comunque verificarsi eventi di esposizione dei dati. Valuta questo aspetto in base alla tolleranza al rischio dei dati sensibili dell'organizzazione: se la fuga di codice sorgente rappresenta un rischio critico, il blocco rigido basato sul rilevamento di pattern di codice vale il disagio iniziale causato dai falsi positivi.

Obiettivi entro i 30 giorni

- Inventario delle applicazioni di intelligenza artificiale con classificazione del livello di rischio (autorizzate, tollerate, proibite)
- Attivazione delle policy di accesso iniziali per le app di intelligenza artificiale e i gruppi di utenti a più alto rischio
- Attivazione della DLP sui prompt per le categorie di dati sensibili con completamento della calibrazione iniziale
- Prima dashboard per la dirigenza con l'impronta AI e le metriche di rischio
- Pubblicazione delle linee guida sull'uso accettabile dell'AI e sulla procedura per la richiesta di eccezioni

Giorni 31–60

Miglioramento della maturità delle policy

Dal giorno 31 al giorno 60, la visibilità e i controlli iniziali vengono consolidati per stabilire un modello di policy maturo, flussi di lavoro di governance funzionanti e un inventario dell'infrastruttura AI.

Checklist per l'implementazione in 60 giorni

Perfeziona il modello della policy di accesso

- **Definisci policy di accesso all'AI specifiche per reparto:** le funzioni di Risorse umane, Finanza, Ingegneria e Vendite hanno esigenze diverse in termini di strumenti AI e presentano profili di rischio differenti per quanto riguarda l'esposizione dei dati
- **Implementa livelli di controllo delle policy condizionali:** livelli di accesso per dispositivi gestiti rispetto a BYOD, segnali di accesso geografico, punteggi di rischio in base al profilo di sicurezza del dispositivo
- **Formalizza tre livelli di accesso per le applicazioni AI:** autorizzato (accesso completo con ispezione inline), tollerato (accesso monitorato con restrizioni d'uso), proibito (bloccato con percorso di eccezione)
- Rivedi e archivia le richieste di eccezione dei primi 30 giorni in base al modello di policy ormai più maturo

Inventario dell'infrastruttura AI

- Esegui l'inventario di modelli, agenti, servizi, dataset, database vettoriali e pipeline AI e crea l'AI-BOM con la tracciabilità della derivazione dei dati, dai dataset fino all'utilizzo del modello in fase di runtime
- **Esegui la valutazione AI-SPM:** identifica configurazioni errate, autorizzazioni eccessive, esposizione di dati sensibili e vulnerabilità nei componenti dell'infrastruttura AI
- **Dai priorità alla correzione dei problemi ad alto rischio relativi ai sistemi di gestione dei processi sensibili basati sull'AI:** autorizzazioni eccessive, dati sensibili esposti e percorsi di accesso non controllati ai componenti AI
- **Convalida la copertura sulle principali piattaforme di AI:** modelli ospitati nel cloud, piattaforme di agenti distribuite a livello aziendale, integrazioni con server MCP e infrastrutture AI non gestite

Definisci processi di governance

- **Associa i requisiti di governance dell'AI alle policy interne e ai framework applicabili:** NIST AI RMF, Legge sull'intelligenza artificiale dell'Unione Europea o HIPAA in caso di informazioni sanitarie protette
- **Formalizza il flusso di lavoro di approvazione degli strumenti AI:** invio, revisione della sicurezza, classificazione del rischio, approvazione condizionale e rivalutazione periodica
- Definisci etichette di sensibilità standard per i tipi di dati rilevanti ai fini dell'utilizzo dell'AI, con regole di gestione chiare per ciascuna etichetta
- **Definisci i percorsi di escalation per gli incidenti di sicurezza relativi all'AI:** chi è responsabile della risposta, quali incidenti devono essere comunicati alla dirigenza e quali evidenze vengono conservate

Proteggi le applicazioni di intelligenza artificiale private

- **Aggiungi il rilevamento dell'iniezione di prompt ai controlli per le app AI private:** input malevoli appositamente elaborati per sovrascrivere le istruzioni del modello o ottenere informazioni sensibili
- **Implementa il monitoraggio del data poisoning per le pipeline RAG:** rileva le modifiche non autorizzate al corpus di documenti che potrebbero falsare gli output del modello
- **Abilita la registrazione di log per prompt/risposte nei sistemi AI privati con il rilevamento delle anomalie:** modelli di query insoliti, contenuti degli output inattesi e interazioni al di fuori dell'ambito previsto
- Definisci flussi di revisione delle interazioni con i modelli per le applicazioni AI private più sensibili

Obiettivi entro i 60 giorni

- Consolidamento del modello della policy di accesso basato sui ruoli con applicazione condizionale in produzione
- Operatività del flusso di governance dell'AI: approvazione degli strumenti, etichettatura della sensibilità, percorsi di escalation definiti
- Completamento della baseline di visibilità AI-SPM con backlog di interventi correttivi prioritari
- AI-BOM, che comprende i principali componenti dell'infrastruttura AI
- Implementazione delle misure di protezione per le applicazioni AI interne con la massima priorità
- Definizione della cadenza dei report sulla conformità e generazione del primo report

Giorni 61–90

Automazione e convalida

Dal giorno 61 al giorno 90, il programma diventa autosufficiente: applicazione automatizzata, monitoraggio continuo della conformità e convalida del sistema con test avversariali tramite il red teaming.

Checklist per l'implementazione in 90 giorni

Applicazione automatizzata

- **Implementa un sistema automatizzato di coaching per gli utenti che violano le norme:** invia avvisi con indicazioni specifiche per il contesto prima di procedere con il blocco
- **Integra il sistema di ticketing ITSM o SOAR per le violazioni ripetute:** crea automaticamente ticket di indagine per gli utenti che attivano ripetutamente gli stessi controlli
- **Configura i trigger di quarantena e isolamento automatici per i comportamenti ad alto rischio:** invio di grandi quantità di dati sensibili, accesso a destinazioni AI proibite, modelli di richiesta anomali
- **Crea circuiti di feedback:** monitora le tendenze di violazione delle norme nel tempo per identificare le categorie di comportamento che suggeriscono la necessità di un adeguamento delle policy piuttosto che il fallimento della loro applicazione

Consolida i report di conformità

- **Configura il monitoraggio sempre attivo per la conformità dell'AI:** quali controlli sono attivi, quali risorse sono coperte e dove esistono lacune nella copertura
- **Associa i controlli ai framework applicabili:** NIST AI RMF (incluso NIST AI 600–1 per il rischio dell'AI generativa), obblighi normativi della Legge sull'intelligenza artificiale dell'Unione Europea, RGPD e HIPAA in caso di informazioni sanitarie protette
- **Crea pacchetti di report pronti per l'audit:** riepilogo dell'utilizzo delle app AI, azioni DLP immediate e risultati, eccezioni concesse e scadute, stato di correzione dei problemi rilevati da AI–SPM
- **Definisci una cadenza trimestrale per la revisione del programma di sicurezza dell'AI:** prestazioni delle policy, lacune nella copertura, nuove aree di rischio e modifiche al piano d'implementazione

Attività di red teaming e rafforzamento della sicurezza

- **Stabilisci un processo periodico di attività di red teaming per le applicazioni AI sviluppate internamente:** librerie preconfigurate di test d'attacco integrate con casi di verifica personalizzati in base al contesto specifico dell'applicazione
- **Esegui test che coprano l'intera superficie di attacco dell'AI:** iniezione di prompt, jailbreak, avvelenamento del contesto, estrazione di informazioni personali, perdita di codice sorgente, comportamento anomalo e regressione del benchmark del modello.
- **Implementa policy di protezione del runtime per i sistemi AI di produzione:** rilevatori di prompt injection, fughe di dati (informazioni personali e codice sorgente), tentativi di jailbreak e violazioni della moderazione dei contenuti.
- **Automatizza la traduzione dei risultati del red teaming in misure di sicurezza del runtime:** il divario tra i test e l'applicazione in fase di produzione è il punto in cui la maggior parte dei programmi di sicurezza delle applicazioni AI fallisce.

16 minuti	72%	100%
Tempo mediano al primo errore critico nei test di red teaming dell'AI. Oltre 222.000 attacchi simulati in più di 25 ambienti	delle aziende ha rivelato una vulnerabilità critica già al primo test. Il rischio critico è presente fin dal primo giorno, anche in ambienti maturi.	dei sistemi AI testati presentava almeno una vulnerabilità critica. Nessun sistema AI è sicuro per impostazione predefinita. I test continui sono obbligatori.

Convalida l'allineamento dell'architettura

- **Mappa tutti i percorsi di accesso all'AI rispetto ai sei principi dello zero trust:** verifica che ciascun principio sia applicato a tutte le superfici di accesso del personale, degli sviluppatori e delle infrastrutture AI.
- **Individua e colma le lacune normative nei punti di intersezione:** punti in cui termina un livello di applicazione delle policy e ne inizia un altro senza una copertura coerente.
- Standardizza l'applicazione delle policy attraverso un modello unificato che comprenda la verifica dell'identità, il controllo delle destinazioni, la valutazione del rischio, l'applicazione inline, la minimizzazione di accesso ed esposizione e il monitoraggio.

Obiettivi entro i 90 giorni

- Operatività dei flussi di lavoro automatizzati per la risposta alle violazioni e il coaching degli utenti
- Documentazione del monitoraggio continuo della conformità in tempo reale con allineamento ai framework
- Pacchetto di reportistica sulla governance destinato alla dirigenza e agli audit
- Definizione di una cadenza per le attività di red teaming AI con policy attive di protezione del runtime
- Conversione dei risultati delle attività di red teaming in misure di protezione operative attraverso un processo a ciclo chiuso
- Scheda di valutazione del programma e piano d'azione trimestrale per la governance continua

In che modo Zscaler favorisce l'adozione sicura dell'intelligenza artificiale

La piattaforma Zero Trust Exchange™ è stata creata per la governance e la protezione dell'intelligenza artificiale su scala aziendale. Si tratta di un'estensione della stessa piattaforma che già gestisce l'accesso di utenti, workload e reti per migliaia di aziende in tutto il mondo, e non di un'implementazione a posteriori su un'architettura di sicurezza preesistente.

Ogni fase del ciclo di miglioramento continuo è associata a un insieme specifico di funzionalità della piattaforma. La tabella seguente mostra cosa offre Zscaler a ciascun livello.

RILEVAMENTO Conosci l'impronta totale dell'AI	CONTROLLO Applica l'accesso zero trust per l'AI	PROTEZIONE Rafforza e proteggi le applicazioni AI
Gestione delle risorse IA Zscaler accede a ogni applicazione, modello, agente, server MCP e funzionalità AI SaaS integrata nel tuo ambiente, inclusi gli strumenti che il reparto IT non ha mai approvato.	**AI Access Security** Applica controlli di accesso zero trust per i servizi AI nelle app SaaS, l'AI integrata nelle piattaforme aziendali e gli ambienti di sviluppo AI, con ispezione inline a ogni livello.	**Attività di red teaming basate sull'AI e vincoli di sicurezza per l'AI** Testa le applicazioni di intelligenza artificiale sviluppate internamente in condizioni reali di attacco, quindi traduci automaticamente i risultati in policy di protezione del runtime.
• Rilevamento della Shadow AI e visibilità sull'utilizzo in tutte le popolazioni di utenti • **AI Bill of Materials (AI-BOM):** tracciabilità dai dataset a modelli, agenti e utilizzo del runtime • **AI-SPM:** valutazione continua del profilo di sicurezza per individuare configurazione errate, autorizzazioni eccessive ed esposizione dei dati sensibili • Rilevamento degli agenti AI negli agenti distribuiti in azienda e integrati nei servizi SaaS	• Criteri granulari di autorizzazione, avviso, isolamento e blocco per gruppo di utenti, applicazione e tipi di dati • DLP inline su prompt, risposte e upload di file su tutto il traffico AI • Zero Trust Browser per l'accesso dei dispositivi non gestiti e BYOD agli strumenti di intelligenza artificiale • Estrazione e classificazione rapide per le principali app di GenAI	• **Red teaming automatizzato e continuo:** iniezione di prompt, jailbreak, avvelenamento del contesto, perdita di dati e variazione del comportamento • Rafforzamento dei prompt e benchmarking dei modelli con mappatura della governance rispetto a NIST AI RMF, Legge sull'intelligenza artificiale dell'Unione Europea, OWASP LLM Top 10 • **Controlli di protezione del runtime:** strumenti di rilevamento di jailbreak, perdita di codice sorgente o informazioni personali, moderazione dei contenuti e comportamenti anomali • **Generazione di policy a ciclo chiuso:** i risultati del red teaming diventano automaticamente strumenti di rilevamento

IL VANTAGGIO DELLA PIATTAFORMA

La maggior parte dei fornitori di soluzioni di sicurezza risolve solo una parte del problema della sicurezza dell'AI. Zscaler copre l'intero ciclo di vita dell'AI su un'unica piattaforma: individua gli strumenti AI ovunque, controlla chi accede a quale intelligenza artificiale utilizzando una policy zero trust e proteggi le applicazioni e l'infrastruttura AI dalla fase di sviluppo all'esecuzione. Il collegamento continuo tra le evidenze del red teaming e i controlli di protezione del runtime in produzione è una funzionalità che nessuna soluzione singola può replicare.

FAQ

Cosa significa "intelligenza artificiale zero trust" nel concreto?
Significa applicare il principio dello zero trust, ovvero non presumere alcuna fiducia implicita ed eseguire verifiche continue a ogni interazione con l'AI nel proprio ambiente. In pratica: nessuna app, utente, dispositivo, richiesta o connettore AI deve essere considerato intrinsecamente affidabile. L'accesso viene consentito in base all'identità verificata, al profilo del dispositivo, alla sensibilità dei dati e al contesto della sessione. I controlli vengono applicati in tempo reale. Ogni decisione di accesso viene registrata ed è verificabile.

Cosa dobbiamo fare nei primi 30 giorni per ridurre il rischio della Shadow AI?
Inizia con la fase di rilevamento. Non si può governare un'AI che non si vede, e ThreatLabz traccia oltre 3.400 applicazioni di AI nel traffico aziendale, un numero che si è quadruplicato in un solo anno, in gran parte a causa delle funzionalità AI integrate negli strumenti già approvati dal reparto IT. Una volta ottenuta la visibilità, classifica le app più rischiose e definisci le policy di destinazione di base. Aggiungi la protezione dei dati (DLP) per le categorie di dati a più alto rischio. Pubblica una guida chiara sull'utilizzo accettabile e una procedura per la richiesta di eccezioni, in modo che gli utenti abbiano un percorso ufficiale da seguire ed evitino di aggirare i controlli.

In che modo i controlli di accesso dell'AI e la DLP inline proteggono i prompt e gli upload?
I controlli di accesso determinano chi può raggiungere quali strumenti AI, da quali dispositivi e posizioni e in quali condizioni. La DLP inline opera a livello di contenuto, ispezionando ciò che gli utenti inviano, il testo della richiesta, l'upload di file e le attività di copia/incolla in tempo reale. I due controlli sono complementari: il controllo degli accessi limita chi può accedere a quale strumento AI, mentre la DLP controlla quali dati transitano attraverso i canali consentiti. Una lacuna in uno qualsiasi di questi due ambiti genera un'esposizione, motivo per cui le organizzazioni che implementano la DLP per gli upload ma non effettuano un'ispezione immediata corrono comunque un rischio significativo di perdita di dati.

Cos'è AI-SPM e perché è importante per l'adozione dell'AI?
AI-SPM è una valutazione continua delle configurazioni, delle autorizzazioni e dei rischi di esposizione dell'infrastruttura AI, l'equivalente per l'AI del CSPM per l'infrastruttura cloud. Individua le risorse AI, ne valuta le configurazioni rispetto alle best practice di sicurezza, rileva le autorizzazioni eccessive e l'esposizione di dati sensibili e monitora le variazioni del profilo di sicurezza nel tempo. È importante perché i rischi dovuti alle configurazioni errate nell'infrastruttura AI sono numerosi e poco compresi dalla maggior parte delle organizzazioni che passano dalla sperimentazione all'implementazione dell'intelligenza artificiale.

Quali indicatori di performance dimostrano meglio alla dirigenza che il rischio legato all'AI sta diminuendo senza bloccare la produttività?
L'approccio più efficace combina una metrica di riduzione del rischio con una metrica relativa all'adozione, in modo che la dirigenza possa osservare entrambe le tendenze contemporaneamente. La riduzione del numero di app AI non autorizzate (Shadow AI) e il contestuale aumento dell'adozione di app AI autorizzate mostrano chiaramente che la governance sta funzionando. Le azioni DLP tempestive (distribuzione delle azioni di blocco, avviso o autorizzazione) mostrano come viene calibrata l'applicazione delle policy. I tempi di risposta alle richieste di eccezione e i tempi del ciclo di approvazione degli strumenti AI dimostrano la reattività della governance.

Agisci rapidamente.
Sicurezza che dura.

Informazioni su Zscaler

Zscaler (NASDAQ: ZS) è pioniera e leader globale nella sicurezza zero trust. Le più grandi aziende del mondo, le organizzazioni che gestiscono infrastrutture critiche e le agenzie governative si affidano a Zscaler per proteggere utenti, filiali, applicazioni, dati e dispositivi e per accelerare le iniziative di trasformazione digitale. Distribuita in oltre 160 data center a livello globale, la piattaforma Zscaler Zero Trust Exchange™, combinata con funzionalità avanzate di intelligenza artificiale, contrasta ogni giorno miliardi di minacce informatiche e violazioni delle policy, consentendo alle imprese moderne di aumentare la produttività riducendo costi e complessità.

+1 408.533.0288 **Zscaler, Inc. (HQ) • 120 Holger Way • San Jose, CA 95134** **zscaler.com/it**