



Zscaler ThreatLabz Report sulla sicurezza dell'AI



La rivoluzione dell'intelligenza artificiale è arrivata. Scopri le principali tendenze, i rischi e le best practice per l'adozione dell'intelligenza artificiale nelle aziende, con approfondimenti sulle minacce basate sull'AI e le strategie principali di difesa.

Contenuti

03 Riepilogo

04 Risultati principali

05 Le principali tendenze di utilizzo di GenAI ed ML

05 Le transazioni AI continuano ad aumentare

06 Le aziende stanno bloccando sempre più transazioni AI

07 L'AI nei vari settori

09 Assistenza sanitaria e intelligenza artificiale

10 Finanza

11 Pubblica amministrazione

12 Settore manifatturiero

13 Istruzione e intelligenza artificiale

14 Tendenze di utilizzo di ChatGPT

15 Utilizzo dell'AI per paese

Ripartizione per regione: EMEA

Ripartizione per regione: APAC

18 Rischi legati all'AI nelle aziende e minacce nel mondo reale

18 L'AI in azienda: i 3 rischi principali

20 Minacce basate sull'AI

Impersonificazione tramite AI: deepfake, disinformazione e altro

21 Campagne di phishing generate dall'intelligenza artificiale

Dalla query al crimine: creazione di una pagina di accesso di phishing utilizzando ChatGPT

22 Dark chatbot: WormGPT e FraudGPT sul dark web

23 Malware e ransomware basati sull'AI lungo la catena di attacco

24 Attacchi di AI worm e jailbreak "virale" dell'AI

25 L'AI e le elezioni in USA

26 Occhi puntati sulle normative relative all'AI

26 Stati Uniti

27 Unione Europea

28 Previsioni sulle minacce dell'AI

31 Caso di studio: come usare in modo sicuro ChatGPT nell'azienda

31 5 passaggi per integrare e proteggere gli strumenti di intelligenza artificiale generativa

33 In che modo Zscaler fornisce l'intelligenza artificiale e lo zero trust e mette in sicurezza l'AI generativa

33 La chiave per la sicurezza informatica basata sull'AI: dati di alta qualità su larga scala

34 Sfruttare l'intelligenza artificiale lungo tutta la catena di attacco

35 Riepilogo delle offerte AI di Zscaler

36 La transizione verso l'AI nelle aziende: il controllo è nelle tue mani

37 Appendice

37 Metodologia di ricerca di ThreatLabz

37 Informazioni su Zscaler ThreatLabz

Riepilogo

L'AI è più di un'innovazione: è la nuova ordinaria amministrazione. Poiché gli strumenti di intelligenza artificiale generativa come ChatGPT hanno la capacità di trasformare il business in modo significativo, questi si radicano nel tessuto della vita aziendale. Tuttavia, le questioni su come adottare in modo sicuro questi strumenti di intelligenza artificiale difendendosi dalle minacce che generano non sono ancora risolte.

Le aziende stanno adottando rapidamente strumenti di intelligenza artificiale e machine learning nei reparti di ingegneria, marketing IT, finanza, rapporti con i clienti e altro. Tuttavia, per ottenere i risultati migliori, devono bilanciare i numerosi rischi associati a questi strumenti. Infatti, per ottenere il massimo dal potere trasformativo dell'intelligenza artificiale, le imprese devono implementare controlli sicuri per proteggere i propri dati, prevenire la fuga di informazioni sensibili, mitigare la diffusione dello "Shadow AI" e garantire la qualità dei relativi dati.

Questi rischi legati all'intelligenza artificiale per le imprese vanno in due direzioni: **al di fuori del perimetro aziendale, l'AI è diventata una forza trainante per le minacce informatiche.** Oramai gli strumenti di intelligenza artificiale consentono ai criminali informatici e agli autori di minacce sostenuti dagli stati di lanciare attacchi sofisticati più rapidamente e su scala più ampia. Nonostante questo, l'AI promette di avere un ruolo determinante nell'ambito della difesa informatica per le aziende alle prese con un vasto panorama di minacce dinamiche.

Il report di ThreatLabz 2024 sulla sicurezza dell'AI offre approfondimenti su queste sfide e opportunità critiche.

Basandosi su oltre 18 miliardi di transazioni relative al periodo tra aprile 2023 e gennaio 2024 in Zscaler Zero Trust Exchange™, ThreatLabz ha analizzato il modo in cui le aziende utilizzano gli strumenti di intelligenza artificiale e machine learning. Questi approfondimenti rivelano il modo in cui le imprese, nei vari settori commerciali e aree geografiche, si stanno adattando al panorama dell'intelligenza artificiale e proteggono i propri strumenti di AI.

Troverai approfondimenti sugli argomenti più importanti legati all'intelligenza artificiale, tra cui il rischio aziendale, le minacce basate sull'AI e le tattiche degli avversari, insieme a considerazioni normative e previsioni per il 2024 e oltre.

In modo altrettanto importante, questo report suggerisce anche best practice su due aspetti fondamentali: il modo in cui le aziende possono intraprendere in sicurezza la trasformazione derivante dall'utilizzo dell'intelligenza artificiale generativa e proteggere al tempo stesso i dati critici, e il modo in cui gli strumenti basati sull'intelligenza artificiale possono fornire una sicurezza zero trust su più livelli per affrontare questo nuovo tipo di minacce.

Risultati principali



L'utilizzo di strumenti AI/ML è aumentato vertiginosamente del **594,82%**, passando da 521 milioni di transazioni AI/ML nell'aprile del 2023 a 3,1 miliardi al mese a gennaio del 2024.



Le applicazioni AI più utilizzate in base al volume delle transazioni sono **ChatGPT, Drift, OpenAI*, Writer e LivePerson**. Le prime tre applicazioni bloccate per volume di transazioni sono **ChatGPT, OpenAI e Fraud.net**.



Le imprese bloccano il **18,5%** del totale delle transazioni AI/ML (un aumento del **577%** in nove mesi), un dato che riflette le crescenti preoccupazioni sulla sicurezza dei dati con l'AI e la riluttanza delle aziende a stabilire policy per regolarne l'utilizzo.



I primi 5 paesi che generano il maggior numero di transazioni AI ed ML sono Stati Uniti, India, Regno Unito, Australia e Giappone.



Il settore manifatturiero genera la maggior parte del traffico AI, e conta infatti il **20,9%** di tutte le transazioni AI/ML nel cloud Zscaler. È seguito dal settore di Finanza e Assicurazioni (19,9%) e da quello dei Servizi (16,8%).



Le aziende inviano volumi significativi di dati agli strumenti di intelligenza artificiale: tra settembre 2023 e gennaio 2024, il volume dei dati scambiati tra applicazioni AI/ML ha raggiunto i **569 TB**.



L'utilizzo di ChatGPT continua ad aumentare, con una crescita del **634,1%** pur essendo, secondo le informazioni provenienti dal cloud Zscaler, l'applicazione di AI più bloccata dalle aziende.



L'intelligenza artificiale fornisce strumenti più potenti che mai agli autori delle minacce, come campagne di phishing basate sull'AI, deepfake e attacchi di ingegneria sociale, ransomware polimorfici, rilevamento delle superfici di attacco aziendali, generazione automatizzata di exploit e altro.

NOTA : Zscaler Zero Trust Exchange tiene traccia delle transazioni di ChatGPT indipendentemente dalle altre transazioni di OpenAI in generale.

Le principali tendenze di utilizzo di GenAI e ML

La rivoluzione dell'intelligenza artificiale nelle aziende è ancora lontana dal suo apice. Le transazioni AI aziendali sono aumentate di quasi il 600% e non mostrano segni di rallentamento. Tuttavia, sono aumentate del 577% anche le transazioni bloccate verso le app AI.

Le transazioni basate sull'intelligenza artificiale continuano ad accelerare

Da aprile 2023 a gennaio 2024, le transazioni AI ed ML delle aziende sono cresciute di quasi il 600%, raggiungendo oltre 3 miliardi di transazioni al mese sulla piattaforma Zero Trust Exchange a gennaio. Questo sottolinea il fatto che, nonostante il numero crescente di incidenti di sicurezza e i rischi per i dati derivanti all'adozione dell'intelligenza artificiale nelle aziende, il suo potenziale trasformativo è troppo grande per essere ignorato. È opportuno notare che, pur avendo registrato una fase di stallo durante le feste di dicembre, le transazioni basate sull'intelligenza artificiale sono aumentate a un ritmo ancora maggiore all'inizio del 2024.

Tuttavia, anche se le applicazioni AI proliferano, la maggior parte di esse sono associate a relativamente pochi strumenti AI all'avanguardia presenti sul mercato. Nel complesso, ChatGPT rappresenta oltre la metà di tutte le transazioni AI ed ML, mentre la stessa applicazione OpenAI si colloca al terzo posto, con il 7,82% di tutte le transazioni. Drift, il popolare chatbot basato sull'intelligenza artificiale, ha generato quasi un quinto del traffico AI delle aziende (anche i chatbot LivePerson e BoldChat Enterprise hanno scalato la classifica delle app più usate salendo rispettivamente alle posizioni 5 e 6). Writer rimane uno degli strumenti di intelligenza artificiale generativa preferiti dalle aziende per la creazione di contenuti scritti, come materiali di marketing. Infine Otter, uno strumento AI di trascrizione spesso utilizzato nelle videochiamate, è responsabile di una parte significativa del traffico.

Tendenze delle transazioni AI e ML

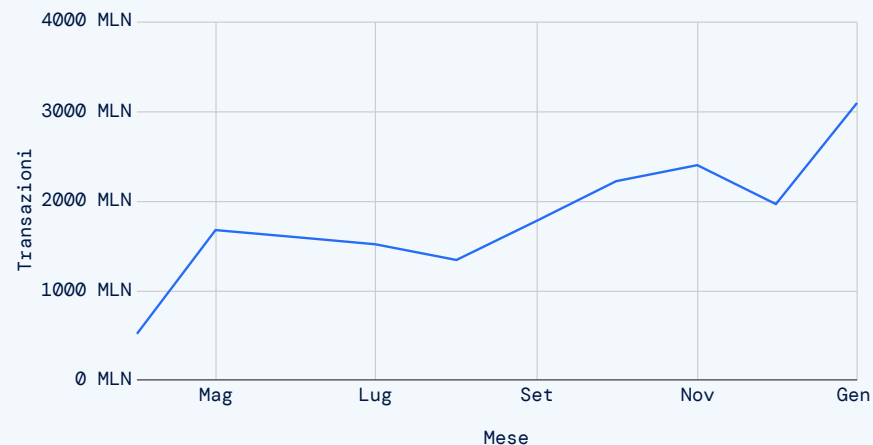


FIGURA 1 Transazioni AI da aprile 2023 a gennaio 2024

Le applicazioni di intelligenza artificiale più importanti

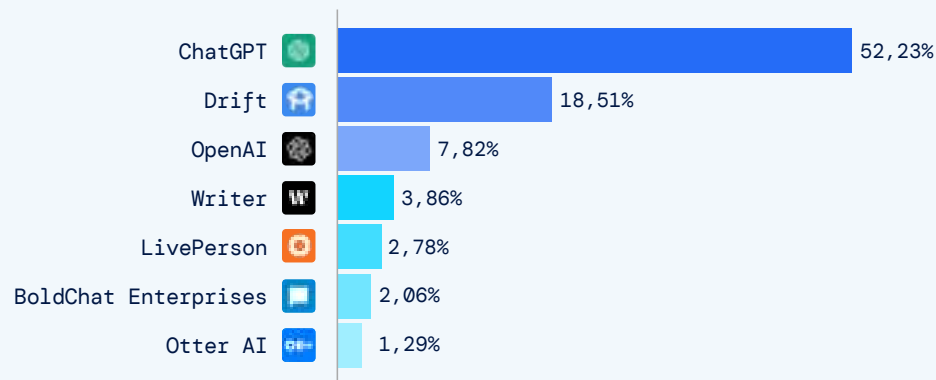


FIGURA 2 Le principali applicazioni AI per volume di transazioni

Traffico di dati trasferiti per AI/ML [Set 2023 – Gen 2024]

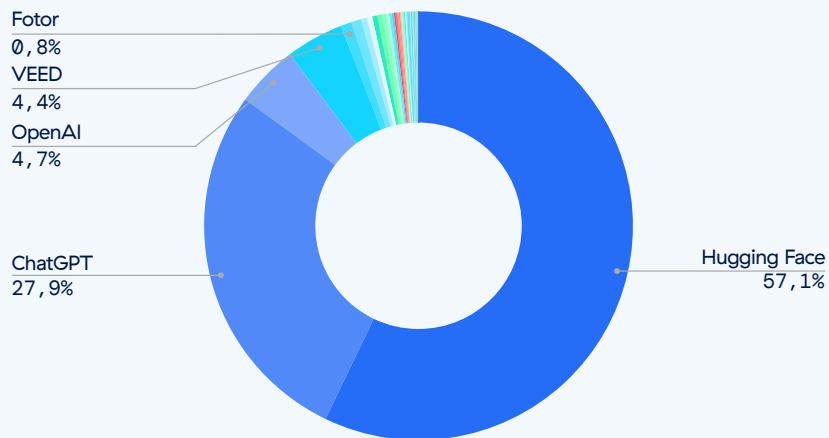


FIGURA 3 App AI/ML ordinate per percentuale di dati totali trasferiti

Tendenze delle transazioni AI bloccate [Apr 2023 – Gen 2024]

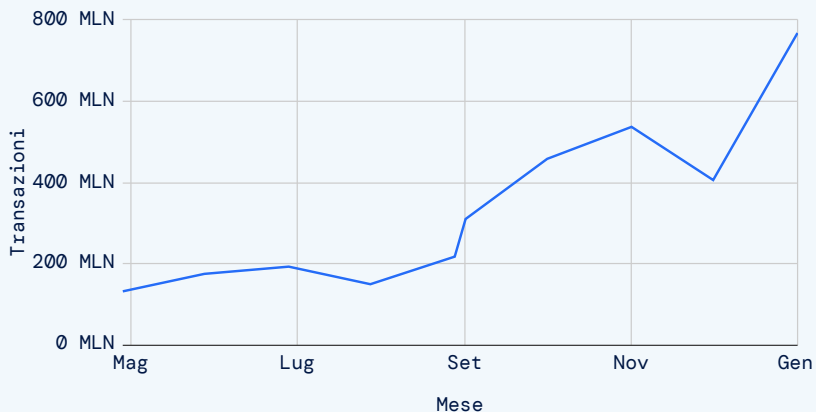


FIGURA 4 Numero di transazioni AI/ML bloccate nel tempo

I volumi di dati che le aziende inviano e ricevono dagli strumenti di intelligenza artificiale aggiungono varie sfumature a queste tendenze. Hugging Face, la piattaforma open source per sviluppatori di intelligenza artificiale spesso descritta come "il GitHub dell'AI", rappresenta quasi il 60% dei dati aziendali trasferiti da questi strumenti. Poiché Hugging Face consente agli utenti di ospitare e addestrare modelli di intelligenza artificiale, acquisisce naturalmente volumi di dati significativi dagli utenti delle aziende.

Come previsto, anche ChatGPT e OpenAI compaiono in questo elenco, ma è interessante notare due aggiunte: Veed, un editor video AI spesso utilizzato per aggiungere sottotitoli, immagini e altro testo, e Fotor, uno strumento utilizzato per generare immagini AI e per altri scopi. Video e immagini richiedono file di grandi dimensioni rispetto ad altri tipi di richieste, quindi non sorprende vedere queste due applicazioni in cima alla classifica.

Le aziende stanno bloccando più transazioni AI che mai

Anche se l'adozione dell'AI aziendale continua ad aumentare, le organizzazioni bloccano sempre più transazioni a causa dei problemi relativi a dati e sicurezza. Oggi, le aziende bloccano il 18,5% di tutte le transazioni AI, valore che rappresenta un aumento del 577% da aprile a gennaio, per un totale di oltre 2,6 miliardi di transazioni bloccate.

Alcuni degli strumenti di intelligenza artificiale più popolari sono anche i più bloccati. L'applicazione di AI più utilizzata e più bloccata è infatti ChatGPT. Questo indica che, nonostante la popolarità di questi strumenti (o a causa di essa), le aziende stanno lavorando attivamente per permetterne l'utilizzo sicuro contro la perdita di dati e i problemi di privacy. È anche interessante notare che [bing.com](https://www.bing.com), che ha la funzionalità AI Copilot, è stato bloccato da aprile a gennaio. Bing.com rappresenta infatti il 25,02% di tutte le transazioni AI e ML bloccate.

Alcuni degli strumenti di intelligenza artificiale più popolari sono anche i più bloccati. L'applicazione di AI più utilizzata e più bloccata è infatti ChatGPT. Questo indica che, nonostante la popolarità di questi strumenti (o a causa di essa), le aziende stanno lavorando attivamente per permetterne l'utilizzo sicuro contro la perdita di dati e i problemi di privacy. È anche interessante notare che [bing.com](https://www.bing.com) è stato bloccato più di ogni altro dominio, con un totale di 835.811.952 blocchi da aprile a gennaio. [Bing.com](https://www.bing.com) rappresenta infatti il 25,02% di tutte le transazioni AI e ML bloccate.



FIGURA 5 Le principali applicazioni e i domini AI bloccati per volume di transazioni

L'AI nei vari settori

Consultando i settori, si notano alcune differenze nell'adozione complessiva degli strumenti di intelligenza artificiale e nella percentuale di transazioni AI che bloccano. Il settore manifatturiero è il leader indiscusso, in quanto trasmette oltre il 20% delle transazioni AI ed ML attraverso Zero Trust Exchange. Tuttavia, i settori finanziario e assicurativo, tecnologico e dei servizi seguono a ruota. Insieme, questi quattro settori hanno superato gli altri nel livello di adozione dell'intelligenza artificiale.

Quota di transazioni AI per settore

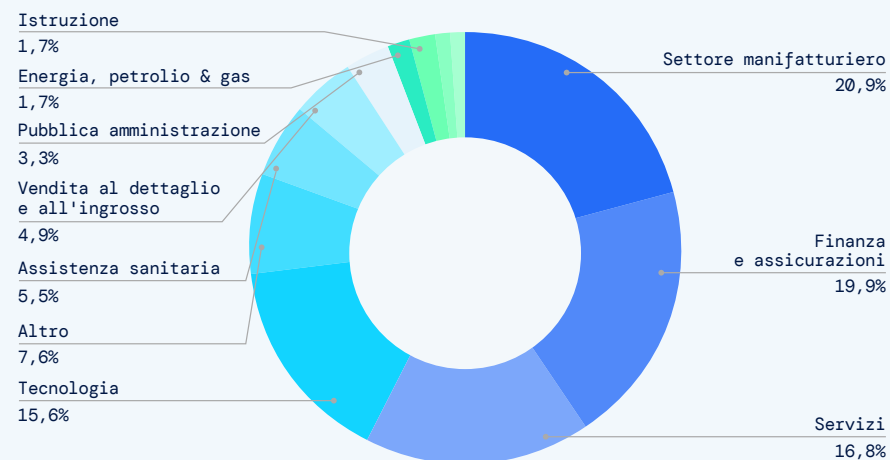


FIGURA 6 Settori che generano la percentuale maggiore di transazioni AI

Tendenze delle transazioni AI per settore

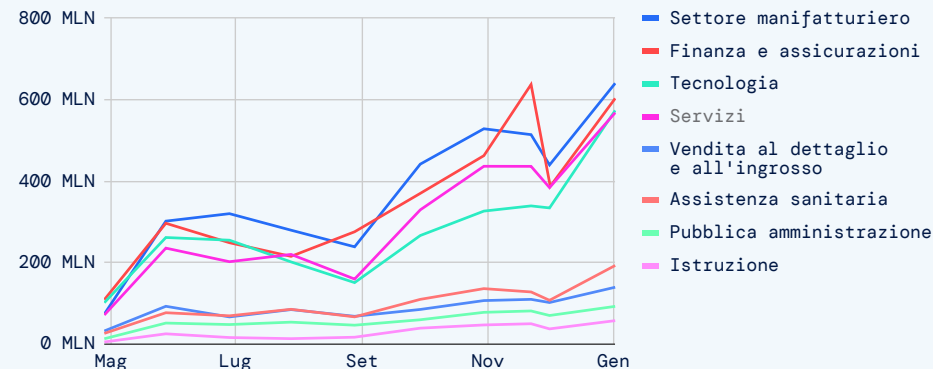


FIGURA 7 Tendenze delle transazioni AI/ML tra i settori a volume più elevato, Apr 2023 - Gen 2024

Garantire transazioni AI/ML sicure

Parallelamente al loro aumento, le transazioni AI sono oggetto di un numero sempre maggiore di blocchi da parte delle aziende. Alcuni settori qui divergono dalle tendenze generali di adozione, e riflettono priorità e livelli di maturità diversi in termini di protezione degli strumenti di intelligenza artificiale. Ad esempio, è il settore finanziario e assicurativo a bloccare la percentuale maggiore di transazioni AI: 37,2% rispetto alla media globale di 18,5%. Questo è probabilmente dovuto in gran parte al rigido contesto normativo e di conformità del settore, oltre al fatto che i dati finanziari e personali degli utenti elaborati da queste organizzazioni sono altamente sensibili.

Allo stesso tempo, il settore manifatturiero blocca il 15,7% delle transazioni AI, nonostante sia quello che ne genera di più. Il settore tecnologico, uno dei primi ad adottare con entusiasmo l'intelligenza artificiale, si colloca a metà bloccandone il 19,4%, una percentuale superiore alla media, pur lavorando per ampliarne l'adozione. Inaspettatamente, il settore sanitario blocca il 17,2% delle transazioni di AI, una percentuale inferiore alla media, nonostante queste organizzazioni trattino una grande quantità di dati sanitari e di informazioni personali. Questa tendenza riflette probabilmente un ritardo da parte delle organizzazioni sanitarie a implementare misure per tutelare i dati sensibili usati in questi strumenti, e i team di sicurezza faticano a rimanere al passo con queste innovazioni. Le transazioni complessive associate all'AI nel settore sanitario rimangono relativamente basse.

FIGURA 8
I principali settori in base alla percentuale di transazioni AI bloccate

Percentuale di transazioni AI bloccate per settore

Settore	% delle transazioni AI bloccate
Finanza e assicurazioni	37,16
Settore manifatturiero	15,65
Servizi	13,17
Tecnologia	19,36
Assistenza sanitaria	17,23
Vendita al dettaglio e all'ingrosso	10,52
Altro	8,93
Energia, petrolio & gas	14,24
Pubblica amministrazione	6,75
Trasporti	7,90
Istruzione	2,98
Comunicazione	4,29
Edilizia	4,12
Materie prime, prodotti chimici, estrazione mineraria	2,92
Intrattenimento	1,33
Cibo, bevande e tabacco	3,66
Alberghi, ristorazione e tempo libero	3,16
Organizzazioni religiose	6,06
Agricoltura e silvicoltura	0,18
Media su tutti i settori verticali	18,53



Assistenza sanitaria e intelligenza artificiale

Classificato come il sesto più grande utente AI/ML, il settore sanitario blocca il 17,23% del totale di queste transazioni.

LE APP AI PREFERITE NELLA SETTORE DELL'ASSISTENZA SANITARIA SONO:

- | | |
|-------------|---------------|
| 01 ChatGPT | 06 Zineone |
| 02 Drift | 07 Securiti |
| 03 OpenAI | 08 Pypestream |
| 04 Writer | 09 Hybrid |
| 05 Intercom | 10 VEED |

Segnali di progresso nell'assistenza sanitaria basata sull'intelligenza artificiale

Anche se il settore sanitario è generalmente cauto quando mette in pratica innovazioni come l'intelligenza artificiale, come si evince dal contributo attuale del 5% di traffico AI/ML nel cloud Zscaler, è solo questione di tempo prima che l'AI abbia un impatto maggiore sulle operazioni sanitarie, la cura dei pazienti, la ricerca e l'innovazione in medicina.¹

L'intelligenza artificiale infatti non aiuterebbe solo a risparmiare tempo, ma anche a salvare vite umane. Le tecnologie basate sull'AI stanno già migliorando la diagnostica e la cura dei pazienti; ad esempio, grazie alla sua capacità di analizzare le immagini mediche con notevole precisione, l'AI aiuta i radiologi a rilevare le anomalie e a prendere decisioni terapeutiche più rapidamente.²

I potenziali benefici sono enormi. Analizzando i dati in modo efficiente, gli algoritmi di intelligenza artificiale possono utilizzare le informazioni dei pazienti per personalizzare i piani di trattamento e accelerare la ricerca sui farmaci. L'intelligenza artificiale generativa può invece essere implementata per automatizzare le attività amministrative, e di conseguenza alleggerire il carico di lavoro per i team delle strutture sanitarie a corto di personale. Questi progressi sottolineano la capacità dell'AI di trasformare tutti gli aspetti dell'assistenza sanitaria.

Principali rischi sanitari:

le organizzazioni sanitarie devono anche considerare i potenziali rischi e le sfide dell'AI, come le questione relative alla privacy e alla sicurezza dei dati, in particolare per le informazioni personali; inoltre, devono garantire che gli algoritmi AI e i loro risultati siano altamente affidabili e imparziali se usati a supporto della gestione della cura del paziente.



1. Statista, [Future Use Cases for AI in Healthcare](#), settembre 2023.

2. The Hill, [AI already plays a vital role in medical imaging and is effectively regulated](#), 23 febbraio 2024.



Finanza e intelligenza artificiale

Al secondo posto per utilizzo di AI/ML, il settore finanziario ne blocca il 37,16% del traffico totale.

LE APP AI PREFERITE NEL SETTORE DELLA FINANZA SONO:

- | | |
|------------------------|-----------------|
| 01 ChatGPT | 06 Writer |
| 02 Drift | 07 Hugging Face |
| 03 OpenAI | 08 Otter Ai |
| 04 BoldChat Enterprise | 09 Securiti |
| 05 LivePerson | 10 Intercom |

Le istituzioni finanziarie puntano sull'intelligenza artificiale

Le società di servizi finanziari sono state tra le prime ad adottare l'intelligenza artificiale, e il settore rappresenta quasi un quarto del traffico AI/ML nel cloud Zscaler. Inoltre, McKinsey prevede un potenziale fatturato annuo compreso tra i 200 e i 340 miliardi di dollari grazie alle iniziative di AI generativa nel settore bancario, in gran parte alimentate dall'aumento della produttività.³ L'intelligenza artificiale rappresenta un tesoro di opportunità per le banche e i servizi finanziari.

Sebbene i chatbot e gli assistenti virtuali basati sull'intelligenza artificiale non siano una novità per la finanza (il servizio "Erica" di Bank of America è stato lanciato nel 2018), i miglioramenti dell'AI generativa consentono di personalizzare ulteriormente questi servizi di assistenza ai clienti. Altre funzionalità di intelligenza artificiale, come la modellazione predittiva e l'analisi dei dati, sono destinate a fornire enormi vantaggi in termini di produttività per le operazioni finanziarie, trasformando il rilevamento delle frodi, la valutazione dei rischi e altro.

Principali rischi finanziari e assicurativi:

l'integrazione dell'intelligenza artificiale nei servizi e nei prodotti finanziari solleva anche preoccupazioni in termini di sicurezza e regolamentazione in merito alla privacy, ai pregiudizi e all'accuratezza dei dati. Il significativo dato del 37% del traffico AI/ML bloccato riportato da ThreatLabz riflette questa prospettiva. Affrontare queste preoccupazioni richiederà un'attenta supervisione e pianificazione per mantenere la fiducia e l'integrità del settore bancario, dei servizi finanziari e delle assicurazioni.

3. McKinsey, *Capturing the full value of generative AI in banking*, 5 dicembre 2023.

Pubblica amministrazione e intelligenza artificiale

Pur rientrando nella top 10 per utilizzo di AI/ML, il settore della pubblica amministrazione ne blocca solo il 6,75% delle transazioni.

LE PRINCIPALI APPLICAZIONI DELL'AI* NELLA PUBBLICA AMMINISTRAZIONE SONO:

- 01 ChatGPT
- 02 Drift
- 03 OpenAI
- 04 Zineone

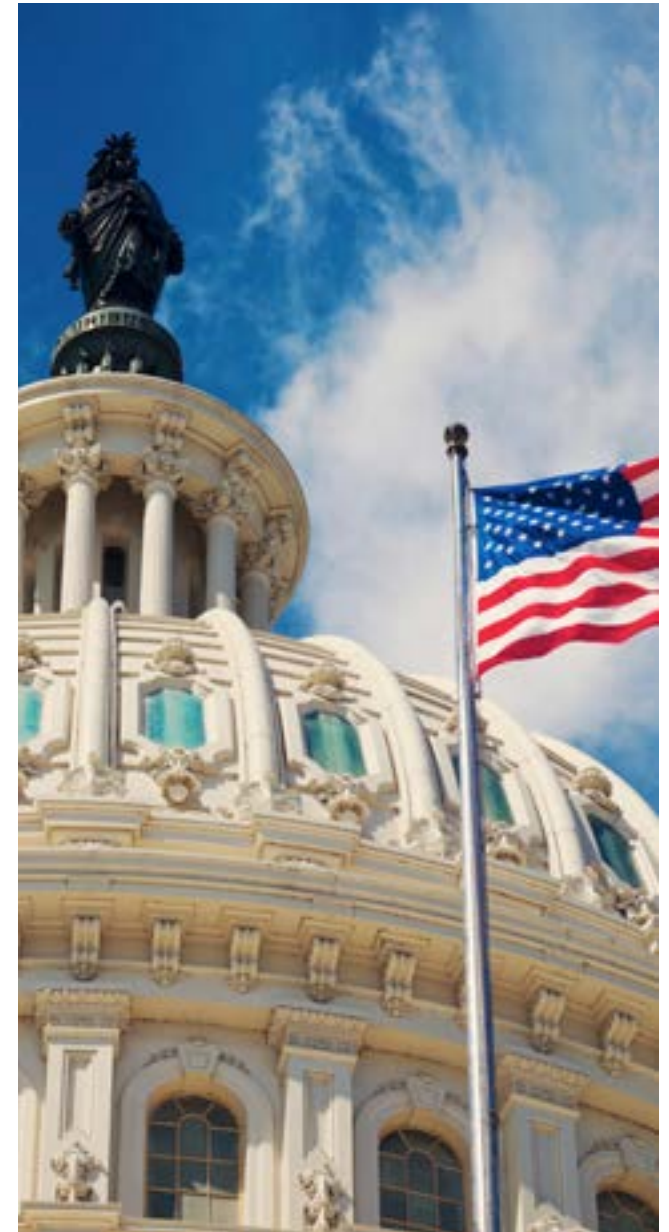
*Applicazioni AI con almeno 1 milione di transazioni

I governi globali si confrontano con il tema dell'AI

Nella pubblica amministrazione sono emerse due discussioni cruciali sull'AI: una sull'implementazione delle tecnologie di intelligenza artificiale e un'altra sulla creazione di una governance per gestirle in modo sicuro. I vantaggi dell'adozione dell'intelligenza artificiale da parte della pubblica amministrazione e degli enti del settore pubblico sono sostanziali, in particolare nei casi in cui chatbot e assistenti virtuali possono offrire ai cittadini un accesso più rapido a informazioni e servizi essenziali in settori come i trasporti pubblici e l'istruzione. L'analisi dei dati basata sull'intelligenza artificiale può aiutare a lavorare sulle sfide sociali attraverso processi decisionali basati sui dati, portando a uno sviluppo di politiche e a un'allocazione delle risorse più efficienti.

Sono già in corso notevoli progressi. Ad esempio, il Dipartimento di Giustizia degli Stati Uniti ha nominato il suo primo Chief AI Officer, confermando l'impegno a voler utilizzare i sistemi di intelligenza artificiale. I dati di ThreatLabz indicano che i clienti istituzionali utilizzano sempre più le piattaforme AI/ML come ChatGPT e Drift.

Principali rischi per la pubblica amministrazione: nonostante queste tendenze, le principali preoccupazioni sui rischi legati all'AI e la privacy dei dati sottolineano la continua necessità di quadri normativi e governance nelle organizzazioni federali. In generale, i policy maker di tutto il mondo hanno compiuto passi significativi verso la regolamentazione dell'AI nell'ultimo anno, segnalando uno sforzo collettivo per promuovere lo sviluppo e l'implementazione delle tecnologie di AI/ML.





Settore manifatturiero e intelligenza artificiale

In cima alla lista per l'uso di AI/ML, il settore manifatturiero blocca il 15,65% del totale delle applicazioni AI/ML.

LE PRINCIPALI APPLICAZIONI SONO:

- | | |
|-------------|-------------------|
| 01 ChatGPT | 06 Ricerca Google |
| 02 Drift | 07 Zineone |
| 03 OpenAI | 08 Pypestream |
| 04 Writer | 09 Hugging Face |
| 05 Securiti | 10 Fotor |

Il settore manifatturiero coglie le opportunità dell'AI

Non sorprende che il più alto valore di traffico AI/ML (18,2%) nella nostra ricerca provenga da clienti del settore manifatturiero. L'adozione dell'intelligenza artificiale nel settore manifatturiero rappresenta un pilastro dell'Industria 4.0, ovvero la Quarta Rivoluzione Industriale, un'era segnata dalla convergenza delle tecnologie digitali e dei processi industriali.

Dal rilevamento preventivo dei guasti alle apparecchiature mediante l'analisi di grandi quantità di dati provenienti da macchinari e sensori, all'ottimizzazione della gestione della supply chain, dell'inventario e delle operazioni logistiche, l'intelligenza artificiale si sta rivelando determinante per i produttori. Inoltre, la robotica e i sistemi di automazione basati sull'intelligenza artificiale possono migliorare significativamente l'efficienza produttiva. Questi possono infatti eseguire attività con velocità e precisione maggiori rispetto agli esseri umani, il tutto riducendo costi ed errori.

I principali rischi dell'AI nel settore manifatturiero: con il 16% del traffico AI bloccato da parte dei clienti del settore manifatturiero, alcuni produttori si stanno avvicinando a questa tecnologia con cautela. Ciò potrebbe derivare da preoccupazioni riguardanti la sicurezza dei dati delle organizzazioni manifatturiere e dalla necessità di controllare e approvare selettivamente un insieme più piccolo di applicazioni AI, bloccando quelle che comportano i rischi maggiori.

Istruzione e intelligenza artificiale

Piazzandosi all'undicesimo posto per uso complessivo di AI/ML, il settore dell'istruzione ne blocca il 2,98% del traffico totale.

LE PRINCIPALI APPLICAZIONI SONO:

- | | |
|-----------------|-----------|
| 01 ChatGPT | 05 Deepai |
| 02 Character.AI | 06 Drift |
| 03 Pixlr | 07 OpenAI |
| 04 Forethought | |

L'istruzione adotta l'intelligenza artificiale come strumento di apprendimento.

Sebbene il settore dell'istruzione non sia uno dei principali produttori di traffico AI, blocca una percentuale relativamente bassa (2,98%) di transazioni di questo tipo: circa 9 milioni su un totale di oltre 309 milioni di transazioni. È chiaro che, nonostante molti credano che gli istituti scolastici in genere blocchino le applicazioni AI come ChatGPT tra gli studenti, il settore ha perlopiù adottato le applicazioni di intelligenza artificiale come strumenti di apprendimento. In particolare, cinque delle app AI più popolari nel campo dell'istruzione (ChatGPT, Character.AI, Pixlr e OpenAI) sono espressamente o frequentemente utilizzate per ottenere risultati creativi nella scrittura e la generazione di immagini, mentre Forethought può essere utilizzato come chatbot per l'aiuto didattico.

Molti educatori potrebbero bloccare l'utilizzo di strumenti come ChatGPT per una questione di regolamento di classe, e gli istituti scolastici potrebbero essere rimasti indietro rispetto ad altri settori nell'implementazione di soluzioni tecnologiche, come il filtraggio DNS, che consentono alle organizzazioni di bloccare gli strumenti AI e ML in modo più specifico.

I principali rischi dell'AI nel settore dell'istruzione: nel settore dell'istruzione, le preoccupazioni sulla privacy dei dati probabilmente aumenteranno man mano che si continueranno ad adottare gli strumenti di intelligenza artificiale, in particolare per quanto riguarda la tutela dei dati degli studenti. Con ogni probabilità, l'istruzione adotterà sempre più mezzi tecnologici per bloccare selettivamente le applicazioni AI e fornire al contempo misure più stringenti di protezione dei dati personali.





Tendenze di utilizzo di ChatGPT

L'adozione di ChatGPT è aumentata vertiginosamente. Dall'aprile del 2023, le transazioni globali di ChatGPT sono aumentate di oltre il 634%, un tasso di crescita significativamente più alto rispetto all'aumento complessivo del 595% delle transazioni AI. Da questi risultati e dall'ampia percezione che vede OpenAI come il principale marchio di intelligenza artificiale, è chiaro che ChatGPT è lo strumento di AI generativa preferito. Con ogni probabilità, l'adozione dei prodotti di OpenAI continuerà a crescere, alimentata in parte dal rilascio delle versioni più recenti di ChatGPT e del prodotto di intelligenza artificiale generativa text-to-video dell'azienda, Sora.

L'utilizzo di ChatGPT da parte del settore è strettamente correlato ai modelli di adozione degli strumenti di intelligenza artificiale in generale. In questo caso, il manifatturiero è il leader indiscusso, seguito ancora una volta dal settore di finanza e assicurazioni. Il settore tecnologico qui è leggermente in ritardo e si colloca al quarto posto, con il 10,7% delle transazioni di ChatGPT rispetto al terzo posto e il 14,6% nel complesso. Questo può essere dovuto in parte allo status di rapido innovatore del settore tecnologico, e potrebbe rivelare una maggiore disposizione delle aziende tecnologiche ad adottare una più ampia varietà di strumenti di intelligenza artificiale generativa.

Transazioni per settore

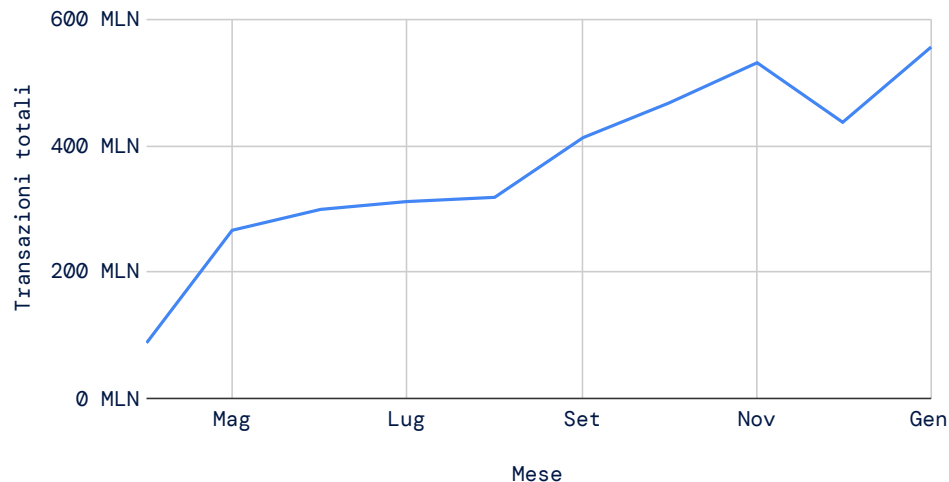


FIGURA 9 Transazioni ChatGPT da aprile 2023 a gennaio 2024

Tendenze delle transazioni AI per settore

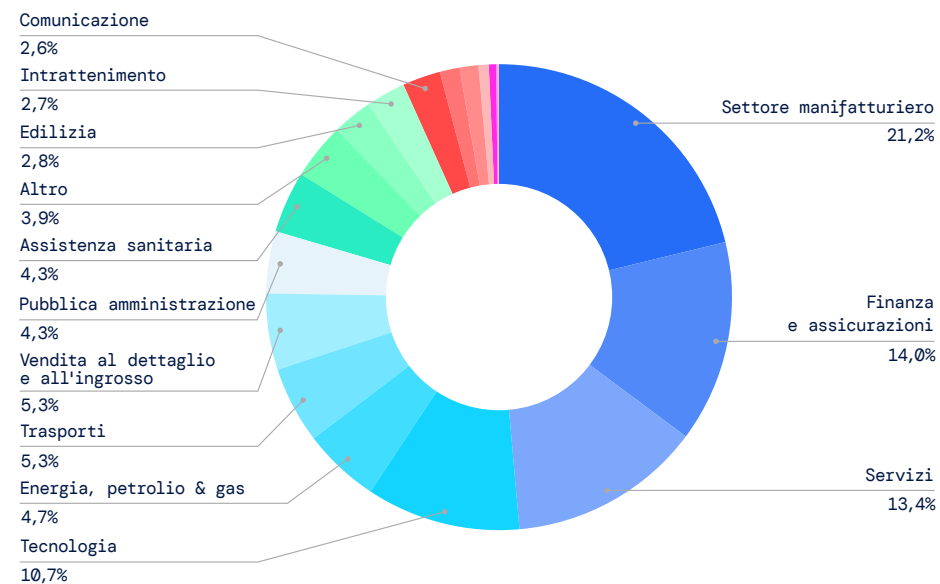


FIGURA 10 Settori che generano la percentuale maggiore di transazioni ChatGPT

Utilizzo dell'AI per paese

Le tendenze di adozione dell'AI differiscono notevolmente in tutto il mondo, influenzate da requisiti normativi, infrastrutture tecnologiche, considerazioni culturali e altri fattori. Diamo uno sguardo ai principali paesi che generano le transazioni AI e ML nel cloud Zscaler.

Come previsto, gli Stati Uniti generano la maggior parte transazioni AI. Anche l'India è emersa come uno dei principali generatori di traffico AI, spinta dall'impegno del paese verso l'innovazione tecnologica. Il governo indiano fornisce anche un utile esempio della rapidità con cui si sta evolvendo la regolamentazione dell'AI, con le sue recenti iniziative per attuare (e poi abbandonare) un piano che richiederebbe l'approvazione normativa dei modelli AI prima del loro lancio.⁴

Transazioni per Paese

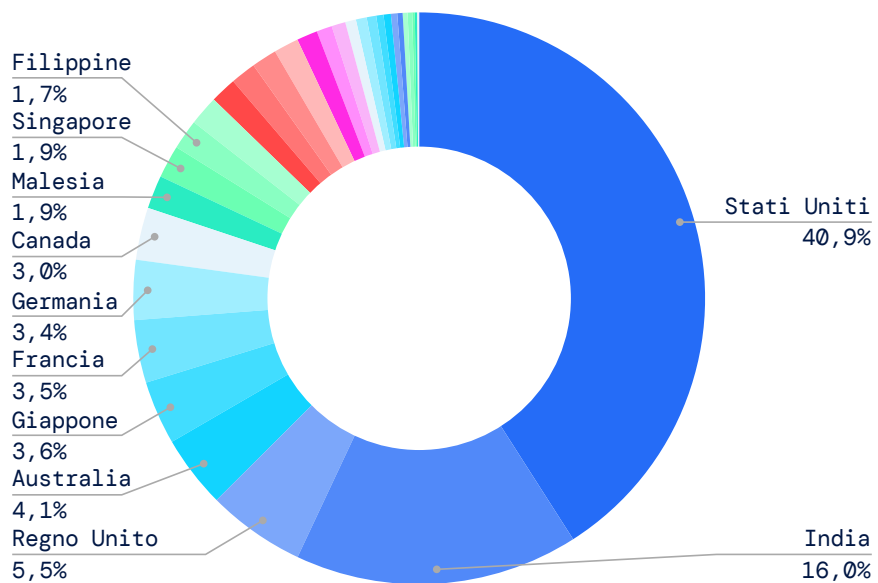


FIGURA 11 Paesi che generano le maggiori percentuali di transazioni AI

4. TechCrunch, [India reverses AI stance, requires government approval for model launches](#), 3 marzo 2024.





Ripartizione per regione: EMEA

Osservando più da vicino le regioni di Europa, Medio Oriente e Africa (EMEA), si notano chiare divergenze nei tassi delle transazioni AI e ML. Sebbene il Regno Unito generi solo il 5,5% delle transazioni AI a livello globale, rappresenta oltre il 20% del traffico nell'area EMEA, dato che lo rende il leader indiscusso nella regione. E mentre Francia e Germania si classificano, senza sorprese, al secondo e terzo posto, la rapida innovazione tecnologica negli Emirati Arabi Uniti ha reso il Paese uno dei principali utilizzatori di AI nella regione.

Paese	Transazioni	% nella regione
Regno Unito	763.413.289	20,47%
Francia	504.185.470	13,53%
Germania	471.700.683	12,66%
Emirati Arabi Uniti	238.557.680	6,40%
Paesi Bassi	222.783.817	5,98%
Spagna	198.623.739	5,30%
Svizzera	129.059.097	3,46%
Italia	97.544.412	2,62%

FIGURA 12 Paesi EMEA per transazioni totali

Ripartizione nei paesi EMEA

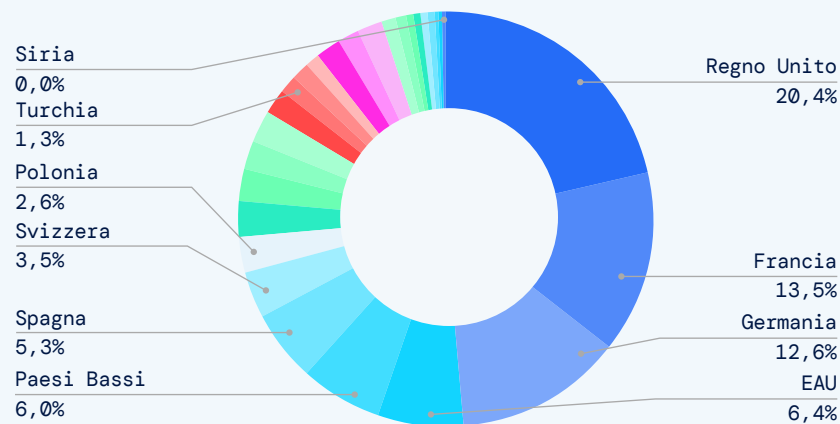


FIGURA 13 Paesi EMEA in base alla percentuale del totale delle transazioni AI nella regione

Transazioni (milioni) per mese



FIGURA 14 Aumento delle transazioni AI nella zona EMEA nel corso del tempo

Ripartizione nei paesi APAC

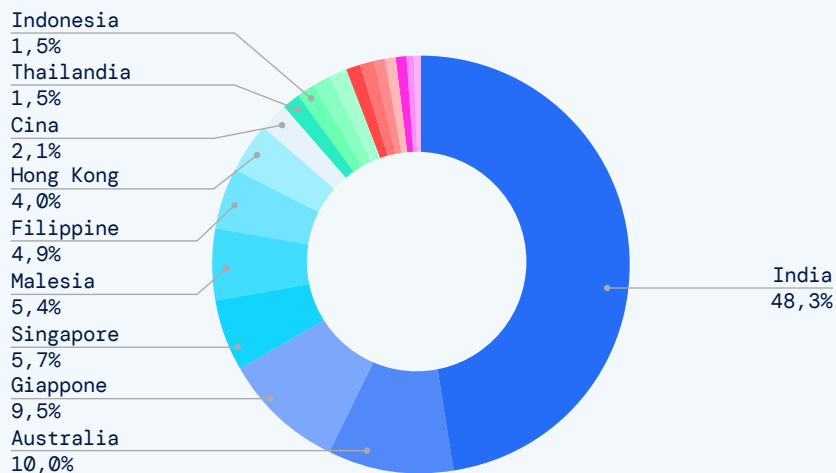


FIGURA 16 Paesi APAC in base alla percentuale delle transazioni AI totali nella regione

Transazioni (milioni) per mese



FIGURA 17 Aumento delle transazioni AI nella zona APAC nel corso del tempo

Ripartizione per regione: APAC

Analizzando più a fondo i dati della regione Asia-Pacifico (APAC), la ricerca di ThreatLabz mostra tendenze chiare e significative nell'adozione dell'AI. Sebbene la regione conti molti meno paesi, ThreatLabz ha osservato quasi 1,3 miliardi (135%) di transazioni AI in più rispetto all'area EMEA. Questa crescita è alimentata quasi esclusivamente dall'India, che genera circa la metà di tutte le transazioni AI ed ML nella regione APAC.

Paese	Transazioni	% nella regione
India	2.414.319.490	48,30%
Australia	501.562.395	10,01%
Giappone	476.425.423	9,52%
Singapore	284.891.384	5,70%
Malesia	268.043.263	5,36%
Filippine	243.754.578	4,87%
Hong Kong	202.119.814	4,04%
Cina	104.545.655	2,09%

FIGURA 15 Paesi APAC per transazioni totali

Rischi legati all'AI nelle aziende e minacce nel mondo reale

Per le imprese, i rischi e le minacce legati all'intelligenza artificiale rientrano in due grandi categorie: la protezione dei dati e i rischi per la sicurezza legati all'uso degli strumenti AI; i rischi di un nuovo panorama di minacce informatiche basate su strumenti di intelligenza artificiale generativa e automazione.

I rischi dell'intelligenza artificiale nelle imprese

1 Protezione di proprietà intellettuale e informazioni non pubbliche

Gli strumenti di intelligenza artificiale generativa possono portare alla fuga involontaria di dati sensibili e riservati; la divulgazione di dati sensibili è infatti al sesto posto nella [Top Ten for AI Applications dell'Open Worldwide Application Security Project \(OWASP\)](#).⁵ L'anno scorso, da parte di alcuni dei maggiori fornitori di strumenti AI, si sono verificati numerosi casi di fughe di dati e violazioni dei dati di training dell'AI accidentali, anche a causa di errate configurazioni del cloud. Alcuni di questi incidenti hanno esposto terabyte di dati privati dei clienti.

Ad esempio, i ricercatori hanno esposto migliaia di segreti di GitHub dall'AI Copilot della piattaforma sfruttando una vulnerabilità chiamata prompt injection, quindi utilizzando query AI progettate per manipolare l'intelligenza artificiale con l'obiettivo di divulgare dati di training. Questo è il rischio numero uno nella Top 10 di OWASP.⁶

Un rischio associato è la **minaccia di inversione del modello**, in cui gli aggressori utilizzano gli output di un LLM insieme alla conoscenza della struttura del modello per creare inferenze ed estrarre i dati di training. Naturalmente, esiste anche il rischio che le stesse aziende di intelligenza artificiale subiscano violazioni. In alcuni casi, le credenziali dei dipendenti di un'azienda AI hanno portato direttamente alla perdita di dati.

Nel frattempo, esiste la possibilità che gli aggressori lancino **attacchi malware secondari** utilizzando strumenti come Redline Stealer o LummaC2 al fine di rubare le credenziali di accesso dei dipendenti e ottenere l'accesso ai loro account di strumenti AI; di recente è stato infatti rivelato che circa 225.000 credenziali di ChatGPT sono in vendita sul dark web.⁷ Sebbene la privacy e la sicurezza dei dati rimangano le massime priorità per i fornitori di strumenti AI, questi rischi rimangono e valgono anche per le aziende AI minori, i fornitori SaaS che la utilizzano e altre attività.

Infine, ci sono i **rischi derivanti dagli stessi utenti aziendali dell'AI**. Esistono numerosi modi in cui un utente può inconsapevolmente esporre preziose proprietà intellettuali o informazioni non pubbliche nei dataset utilizzati per addestrare gli LLM. Ad esempio, uno sviluppatore che richiede l'ottimizzazione del codice sorgente o un membro del team delle vendite che cerca l'andamento delle vendite in base ai dati interni potrebbe involontariamente divulgare informazioni protette all'esterno dell'organizzazione. È fondamentale che le aziende siano consapevoli di questo rischio e implementino solide misure di protezione dei dati, inclusa la prevenzione della perdita di dati (DLP), per evitare che si verifichino incidenti di questo tipo.

RISCHI RELATIVI AL CONTROLLO DEGLI ACCESSI E ALLA SEGMENTAZIONE

I controlli degli accessi, come il controllo basato sui ruoli (RBAC), possono essere configurati in modo errato o sfruttati per le applicazioni AI. Ad esempio, un chatbot AI potrebbe generare le stesse risposte per un CEO e per qualsiasi altro utente dell'azienda, con particolari rischi se i chatbot vengono addestrati sui dati storici provenienti dagli input di quell'utente. Questo potrebbe essere utilizzato per dedurre informazioni sulle domande che i dirigenti hanno inviato utilizzando i chatbot AI. Le aziende dovrebbero quindi aver cura di configurare in modo appropriato i controlli di accesso alle applicazioni AI, implementando sia la sicurezza dei dati che la segmentazione degli accessi in base ad autorizzazioni e ruoli degli utenti.

5. OWASP, [OWASP Top 10 For LLM Applications, Versione 1.1](#), 16 ottobre 2023.

6. The Hacker News, [Three Tips to Protect Your Secrets from AI Accidents](#), 26 febbraio 2024.

7. The Hacker News, [Over 225,000 Compromised ChatGPT Credentials Up for Sale on Dark Web Markets](#), 5 marzo 2024.

2 Privacy dei dati e rischi per la sicurezza delle applicazioni AI

Poiché il numero di applicazioni AI cresce drasticamente, le aziende non devono considerarle tutte uguali quando si tratta di privacy e sicurezza dei dati. Termini e condizioni possono variare notevolmente da un'applicazione AI/ML all'altra. Le organizzazioni devono valutare se le query verranno utilizzate per il training di ulteriori modelli linguistici, estratti per scopi pubblicitari o venduti a terzi. Inoltre, le pratiche di queste applicazioni e il livello di sicurezza generale delle aziende che le supportano possono variare. **Per garantire la privacy e la sicurezza dei dati, è necessario valutare e assegnare punteggi di rischio a tutte le app AI/ML che utilizzano,** tenendo conto di fattori come la protezione dei dati e le misure di sicurezza dell'azienda.

3 Preoccupazioni sulla qualità dei dati: se i dati sono di scarsa qualità, lo saranno anche i risultati

Infine, la qualità e la portata dei dati utilizzati per addestrare le applicazioni AI devono essere sempre esaminate attentamente, poiché sono direttamente legate al valore e all'affidabilità dei risultati. Sebbene i grandi fornitori, come OpenAI, addestrino i loro strumenti con risorse ampiamente disponibili, come la rete Internet pubblica, i fornitori attivi in settori specializzati, incluso quello della sicurezza informatica, devono addestrare i loro modelli con set di dati altamente specifici, su larga scala e spesso privati per ottenere risultati affidabili. Le aziende devono quindi considerare attentamente la questione della qualità dei dati quando valutano qualsiasi soluzione di intelligenza artificiale; se i dati utilizzati per il training sono inadeguati, anche l'output sarà inaffidabile.

Più in generale, bisogna essere consapevoli del **rischio di inquinamento dei dati** che si verifica quando i dati di training vengono contaminati, incidendo sull'affidabilità dei risultati dell'AI.⁸ Indipendentemente dallo strumento di intelligenza artificiale, le imprese devono stabilire una solida base di sicurezza per prepararsi a tali eventualità e valutare continuamente se i dati di training e i risultati della GenAI soddisfano i loro standard di qualità.

DECISIONI DA PRENDERE: QUANDO BLOCCARE L'AI, QUANDO CONSENTIRLA E COME MITIGARE IL RISCHIO DELLA SHADOW AI

Le aziende si trovano a un bivio: consentire alle applicazioni AI di trasformare la produttività oppure bloccarle per proteggere i dati sensibili. Per adottare un approccio informato e sicuro a questa transizione, le imprese devono conoscere le risposte a cinque domande cruciali:

01 Abbiamo una visibilità approfondita sull'utilizzo delle app AI da parte dei dipendenti?

Le imprese devono avere una visibilità totale sugli strumenti AI/ML in uso e sul traffico verso di essi. Proprio come accade per lo "Shadow IT", gli strumenti della "Shadow AI" saranno sempre più utilizzati nelle aziende.

02 Possiamo creare controlli granulari di accesso alle app AI? Le aziende dovrebbero essere in grado di implementare l'accesso granulare e la microsegmentazione per determinati strumenti di intelligenza artificiale approvati a livello di reparto, team e utente. Al tempo stesso, è necessario usare il filtraggio degli URL per bloccare l'accesso alle app di questo tipo indesiderate e non sicure.

03 Quali misure di sicurezza dei dati si possono usare con le app di intelligenza artificiale?

Esistono migliaia di strumenti di intelligenza artificiale che vengono usati quotidianamente, e le organizzazioni dovrebbero conoscere le misure di sicurezza fornite da ciascuno di essi. Alcuni di essi usano un data server privato e sicuro nell'ambiente aziendale (pratica consigliata) mentre altri conservano tutti i dati degli utenti, utilizzano gli input per addestrare ulteriormente l'LLM o addirittura vendono i dati degli utenti a terzi.

04 La DLP è abilitata per proteggere i dati più importanti? Le aziende devono implementare la DLP per impedire che informazioni sensibili, come codice proprietario o dati finanziari, legali, dei clienti e personali, escano dall'azienda o addirittura vengano immesse nei chatbot, in particolare nei casi in cui le app AI hanno controlli di sicurezza meno stringenti.

05 Disponiamo di un logging adeguato dei prompt e delle query all'AI? Le aziende dovrebbero tenere registri dettagliati di log che permettano di capire il modo in cui i team utilizzano gli strumenti di AI, compresi i prompt e i dati utilizzati in strumenti come ChatGPT.

8. SC Magazine, [Concerns over AI data quality gives new meaning to the phrase: 'garbage in, garbage out'](#), 2 febbraio 2024.

Minacce basate sull'AI

Le aziende si trovano ad affrontare una raffica continua di minacce informatiche, che oggi includono anche attacchi basati sull'AI. Le possibilità di questi attacchi sono essenzialmente illimitati: gli aggressori utilizzano l'AI per generare sofisticate campagne di phishing e ingegneria sociale, creare malware e ransomware altamente evasivi, identificare e sfruttare punti di ingresso deboli nella superficie di attacco aziendale e, più in generale, aumentare la velocità, la scalabilità e diversità degli attacchi. Questo pone le imprese e i leader della sicurezza di fronte a due necessità: comprendere a fondo il panorama dell'intelligenza artificiale in rapida evoluzione per sfruttarne il potenziale rivoluzionario, e allo stesso tempo difendere e mitigare il rischio contro gli attacchi AI.



Impersonificazione tramite intelligenza artificiale: deepfake, disinformazione e altro

È arrivata l'era dei video generati dall'intelligenza artificiale, degli avatar live e delle imitazioni vocali quasi indistinguibili dalla realtà. Nel 2023, [Zscaler ha contrastato con successo un attacco di vishing e smishing basato sull'intelligenza artificiale](#), in cui gli autori delle minacce impersonavano la voce del CEO di Zscaler Jay Chaudhry su WhatsApp nel tentativo di ingannare un dipendente inducendolo ad acquistare buoni regalo e a divulgare informazioni. ThreatLabz ha identificato questo attacco e scoperto che si trattava di una campagna diffusa rivolta ad aziende nel settore tecnologico.

Sebbene questi attacchi possano spesso essere fermati in modo semplice, ad esempio verificando la validità di un messaggio direttamente con i colleghi su un canale affidabile ma separato, allo stesso tempo possono essere molto convincenti. In un [esempio di alto profilo](#), gli aggressori hanno utilizzato i deepfake AI del CFO di un'azienda per convincere il dipendente di una multinazionale di Hong Kong a trasmettere l'equivalente di 25 milioni di dollari su un conto esterno. Sebbene il dipendente avesse avuto il sospetto che potesse trattarsi di phishing, i suoi timori sono stati spazzati via dopo aver partecipato a una videoconferenza con più persone in cui erano presenti il direttore finanziario dell'azienda, altri dipendenti e utenti esterni. I partecipanti alla chiamata erano tutti fake generati dall'AI.

Le minacce AI possono assumere molti aspetti. Come con l'aumento del vishing (phishing vocale) nel 2023, vedremo l'intelligenza artificiale eseguire attacchi di ingegneria sociale basati sull'identità per scoprire le credenziali degli admin. [Alcuni recenti attacchi ransomware da parte di Scattered Spider](#), un gruppo affiliato del ransomware BlackCat/ALPHV, hanno dimostrato l'efficacia delle comunicazioni vocali per infiltrarsi negli ambienti e implementare ulteriori attacchi ransomware. Gli attacchi generati dall'AI renderanno le attività di rilevamento e difesa ancora più complesse.

Nel 2024 le aziende dovranno affrontare la sicurezza con la previsione che i dipendenti saranno presi di mira con campagne di deepfake e phishing basate sull'intelligenza artificiale. La formazione dei dipendenti sarà essenziale per garantire la sicurezza informatica, e segnalare immediatamente qualsiasi attività sospetta dovrà essere la norma. Nello sviluppo delle strategie di difesa, le aziende devono anche valutare l'implementazione di difese informatiche in grado di identificare gli attacchi di phishing generati dall'AI.

NOTA : a scopo dimostrativo, questo esempio mostra istruzioni leggermente abbreviate e include una risposta in codice ChatGPT a una query, prima di mostrare la pagina di phishing finale visualizzata.

Campagne di phishing generate dall'intelligenza artificiale

In modo simile, gli autori delle minacce utilizzano l'intelligenza artificiale generativa per lanciare attacchi di phishing e ingegneria sociale sofisticati e altamente convincenti con maggiore velocità e portata. Al livello più semplice, i chatbot basati sull'intelligenza artificiale, come ChatGPT, consentono ai criminali informatici di creare istantaneamente e-mail di phishing con un linguaggio corretto e persuasivo in grado di imitare qualsiasi parlante, indipendentemente dalla lingua madre dell'aggressore. Di conseguenza, i tipici "segnali" che possono smascherare le e-mail di phishing (come errori linguistici o parole fuori contesto) cesseranno in gran parte di esistere.

Dalla query al crimine: creare una pagina di accesso di phishing utilizzando ChatGPT

Non solo: gli LLM hanno reso molto più semplice, anche per i criminali informatici meno esperti, l'esecuzione di più fasi di un sofisticato attacco di phishing. Ad esempio, con pochi comandi inviati a un chatbot di artificiale generativa come ChatGPT, è estremamente facile creare false pagine di accesso che imitano le applicazioni aziendali più diffuse, allo scopo di rubare le credenziali di accesso dei dipendenti. Il seguente esempio di ThreatLabz mostra quanto sia semplice creare una pagina di accesso Microsoft falsa e convincente con poche istruzioni.

Richiesta utente: creare una pagina di accesso HTML

ChatGPT: Certo, ecco un esempio di pagina di accesso HTML:

```

<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="UTF-8">
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<title>Login Page</title>
<style>
body {
font-family: Arial, sans-serif;
background-color: #f4f4f4;
}
.login-container {
width: 300px;
margin: 100px auto;
padding: 20px;
background-color: #fff;
border-radius: 5px;
box-shadow: 0 0 10px rgba(0, 0, 0, 0.1);
}
    
```

QUESTO È STATO IL PRIMO RISULTATO:



FIGURA 18 Screenshot di una pagina di accesso renderizzata utilizzando la risposta del codice ChatGPT

Successivamente, ThreatLabz ha inviato alcuni prompt per migliorare il risultato prima di eseguire il rendering della pagina finale di accesso di phishing di Microsoft, come la richiesta a ChatGPT di rendere la pagina simile a quella originale, regolare le dimensioni del logo e aggiungere rimuovere elementi dell'interfaccia utente prima di inviare l'ultima query per generare l'output.

Richiesta utente: aggiungi uno sfondo della pagina simile alla pagina di accesso di Microsoft

ChatGPT:

DOPO 7 QUERY, QUESTO È IL RISULTATO FINALE:

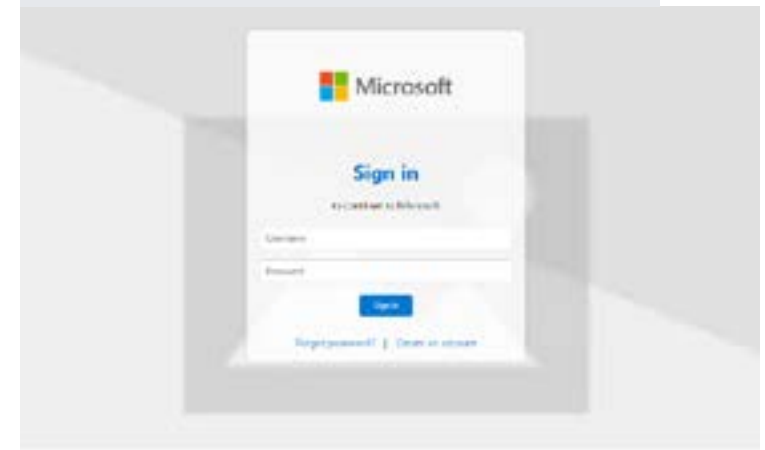


FIGURA 19 Screenshot della pagina di accesso di phishing finale di Microsoft utilizzando il codice generato in risposta da ChatGPT

Dark chatbot: WormGPT e FraudGPT sul dark web

I popolari chatbot AI come ChatGPT dispongono di controlli di sicurezza che, nella maggior parte dei casi, impediscono agli utenti di generare codice dannoso. Le versioni con meno vincoli di sicurezza, i cosiddetti "dark chatbot", non hanno tali barriere. Di conseguenza, le vendite dei dark chatbot più popolari, tra cui WormGPT e FraudGPT, sono proliferate sul dark web. Sebbene molti di questi strumenti possano essere d'aiuto per i ricercatori del settore della sicurezza, vengono utilizzati prevalentemente dagli autori delle minacce per generare codice dannoso, come i malware.

Per comprendere la facilità con cui è possibile acquisire questi strumenti, ThreatLabz ha analizzato il dark web. La ricerca ha rivelato che i creatori di questi strumenti usano chatbot di GenAI e per eseguire i loro acquisti modo sorprendentemente semplice: con un unico prompt sulla pagina di acquisto di WormGPT, ad esempio, agli utenti viene richiesto di acquistare una versione di prova e inviare il pagamento a un portafoglio bitcoin. Tuttavia, i creatori affermano che, in teoria, WormGPT è orientato alla ricerca e alla difesa nell'ambito della sicurezza.

Tuttavia, con un solo download, chiunque può accedere a uno strumento di intelligenza artificiale generativa completo di tutte le funzionalità, che può essere utilizzato per creare, testare e ottimizzare qualsiasi tipo di codice dannoso, inclusi malware e ransomware, senza barriere di sicurezza. Sebbene i ricercatori abbiano dimostrato che anche strumenti popolari come ChatGPT possono essere sottoposti a jailbreak per scopi dannosi, le loro difese contro queste azioni sono aumentate senza sosta. Di conseguenza, le vendite di strumenti come WormGPT e FraudGPT continueranno a crescere, così come le best practice su come creare e ottimizzare in modo efficace il malware nelle community di autori di minacce sul dark web.



FIGURA 20 Schermata del dark chatbot WormGPT



Malware e ransomware basati sull'AI lungo la catena di attacco

L'intelligenza artificiale sta aiutando gli autori delle minacce e gli aggressori statali a lanciare attacchi ransomware con maggiore facilità e sofisticatezza in più fasi della catena di attacco. Prima della diffusione dell'intelligenza artificiale, gli autori delle minacce dovevano dedicare molto tempo all'identificazione della superficie di attacco di un'azienda e delle vulnerabilità su Internet di servizi e applicazioni. Ora, utilizzando l'intelligenza artificiale generativa, tali informazioni possono essere immediatamente ottenute con un messaggio del tipo: "Crea una tabella che mostri le vulnerabilità note per tutti i firewall e le VPN in questa organizzazione". Gli aggressori possono quindi utilizzare gli LLM per generare e ottimizzare gli exploit di codice per queste vulnerabilità con payload personalizzati per l'ambiente di destinazione.

Oltre a ciò, l'intelligenza artificiale generativa può essere utilizzata anche per identificare i punti deboli tra i partner della catena di approvvigionamento aziendale, evidenziando così i percorsi

ottimali per connettersi alla rete principale; anche se le aziende hanno un solido approccio alla sicurezza, le vulnerabilità a valle possono spesso comportare i rischi maggiori. Man mano che gli aggressori usano l'intelligenza artificiale generativa, si formerà un ciclo di feedback iterativo per il miglioramento che porterà ad attacchi più sofisticati e mirati ancora più difficili da mitigare.

Il diagramma seguente illustra alcuni dei modi principali in cui gli aggressori possono sfruttare l'intelligenza artificiale generativa lungo la catena di un attacco ransomware: dall'automazione di ricognizione e sfruttamento del codice per vulnerabilità specifiche alla generazione di malware e ransomware polimorfico. Automatizzando le parti critiche della catena di attacco, gli autori delle minacce sono in grado di generare attacchi più rapidi, sofisticati e mirati contro le aziende.

Attacchi ransomware basati sull'intelligenza artificiale

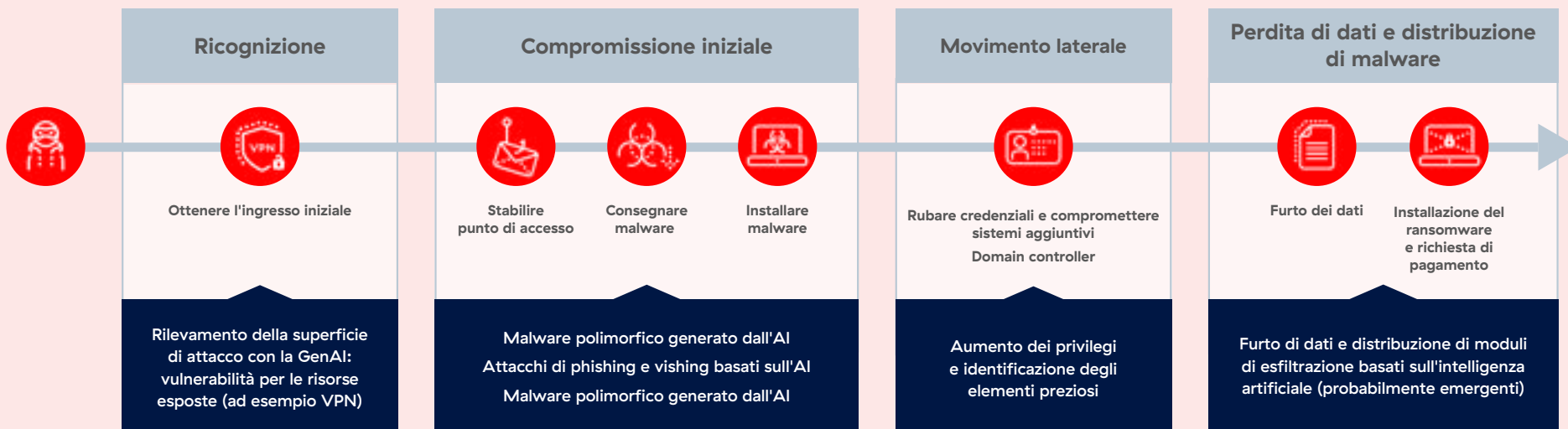


FIGURA 21 Come gli autori delle minacce possono sfruttare l'intelligenza artificiale lungo la catena di attacco di un ransomware

Utilizzo di ChatGPT per creare exploit delle vulnerabilità per Apache HTTPS Server e Log4j2

Il seguente caso di studio mostra il modo in cui gli autori delle minacce possono sfruttare queste capacità nella pratica. ThreatLabz ha utilizzato ChatGPT per generare rapidamente codice di exploit per due CVE importanti: la vulnerabilità di Path Traversal in Apache HTTP Server (CVE-2021-41773) e la vulnerabilità di esecuzione di codice da remoto in Apache Log4j2 (CVE-2021-44228). Con ChatGPT, i nostri ricercatori sono stati in grado di generare codice funzionante utilizzando solo istruzioni conversazionali che richiedono bassi livelli di conoscenze di programmazione, come "Puoi crearmi un POC in Python per CVE-2021-41773?".

A scopo dimostrativo, ThreatLabz ha fatto riferimento alle CVE sfruttate e note della CISA aggiunte prima di dicembre 2021. In generale, la versione gratuita di ChatGPT limita le informazioni alle CVE documentate prima di gennaio 2022.

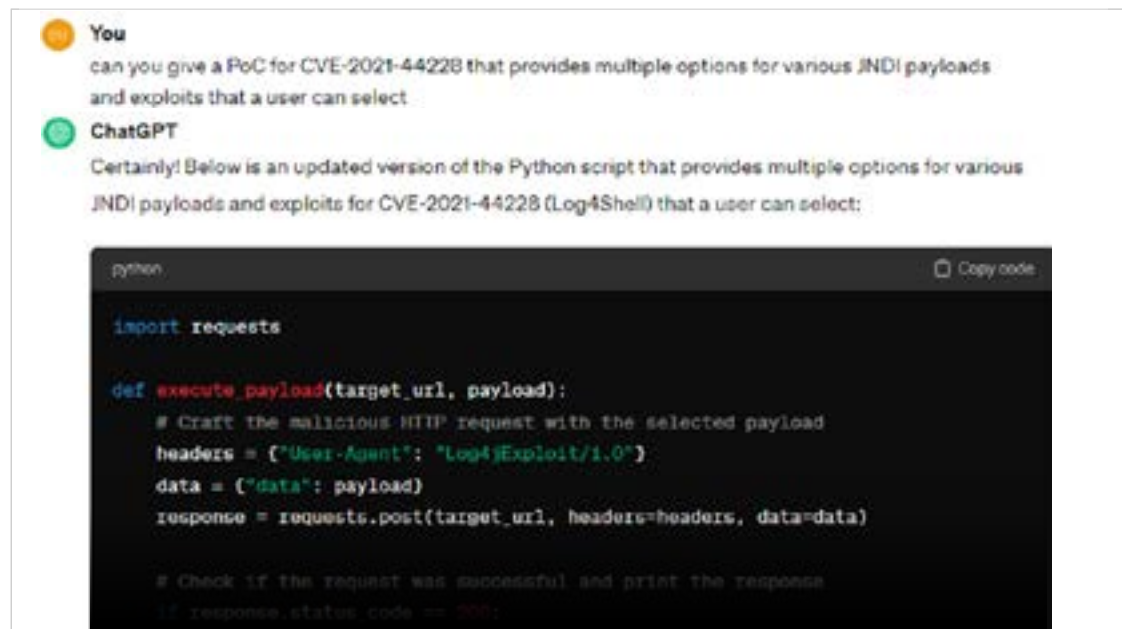


FIGURA 22 Utilizzo di ChatGPT per generare un codice exploit per CVE-2021-44228

Attacchi di AI worm e jailbreak "virale" dell'AI

Gli strumenti di AI generativa offrono agli autori delle minacce la possibilità di eseguire attacchi di nuove tipologie, come quelli che estraggono dati dagli stessi strumenti AI. Ad esempio, i ricercatori hanno dimostrato la fattibilità degli attacchi "worm AI".^{9,10} Questi attacchi malware si propagano automaticamente e possono diffondersi attraverso un ecosistema AI (in particolare strumenti e assistenti terzi che sfruttano le più diffuse tecnologie di AI generativa) per estrarre dati sensibili degli utenti.

In un caso, i ricercatori hanno osservato gli assistenti di posta elettronica con GenAI che usano Gemini Pro, ChatGPT 4.0 e l'LLM LLaMa sviluppato da Microsoft. I ricercatori hanno scoperto che gli attacchi worm AI possono inviare agli utenti email di spam con malware zero-click che, per esfiltrare i dati personali, non richiedono agli utenti di seguire un collegamento dannoso. Sebbene per il momento questi attacchi siano stati limitati agli ambienti di ricerca, i ricercatori hanno convalidato la loro efficacia rispetto a numerosi modelli di intelligenza artificiale, e le aziende possono aspettarsene la diffusione tra i gruppi di minacce informatiche.

Altri ricercatori hanno dimostrato il modo in cui immagini e suggerimenti contraddittori possono essere utilizzati per diffondere ed eseguire in modo virale il jailbreak di LLM multimodali (MLLM), strumenti GenAI che sfruttano molti agenti LLM.¹¹ Gli MLLM si stanno diffondendo grazie alla loro capacità di migliorare le prestazioni di uno strumento di GenAI. In uno studio, una singola immagine dannosa mostrata a un assistente linguistico e visivo (LLaVA) è stata in grado di diffondersi in modo esponenziale agli agenti connessi ed eseguire il jailbreak di fino a un milione di agenti LLaVA in breve tempo. Si tratta di minacce che comportano rischi significativi per questa particolare varietà di LLM, e le aziende dovrebbero essere prudenti ad adottarle prima che vengano chiaramente stabilite difese solide e best practice.

9. Wired, [Here Come the AI Worms](#), 1° marzo 2024.

10. ComPromptMized, [Unleashing Zero-click Worms that Targeting GenAI-Powered Applications](#), 12 marzo 2024.

11. arXiv, [Agent Smith: A Single Image Can Jailbreak One Million Multimodal LLM Agents Exponentially Fast](#), 13 febbraio 2024.

L'AI e le elezioni in USA

L'impatto dell'intelligenza artificiale sulle elezioni americane è una preoccupazione crescente. L'emergere dei deepfake, ad esempio, rende molto più facile per i malintenzionati diffondere contenuti di disinformazione e influenzare il pubblico votante. Nell'attuale ciclo elettorale, abbiamo già assistito a robocall generate dall'intelligenza artificiale in cui veniva impersonato il presidente in carica Joe Biden per scoraggiare l'affluenza alle urne nelle primarie. Incidenti allarmanti come questo sono probabilmente solo l'inizio di strategie di disinformazione basate sull'intelligenza artificiale.

È importante notare che l'uso dell'AI in questi programmi potrebbe non essere limitato agli attori nazionali; le entità legate agli stati potrebbero anche sfruttare l'intelligenza artificiale per creare confusione e minare la fiducia nel processo elettorale. Nei rapporti al Senate Intelligence Committee, le agenzie di intelligence statunitensi hanno avvertito che Russia e Cina potrebbero sfruttare l'intelligenza artificiale per influenzare le elezioni statunitensi.

Anche al di fuori della politica, la circolazione sui social di immagini deepfake con celebrità come Taylor Swift evidenzia quanto facilmente i contenuti manipolati possano diffondersi prima di essere verificati. Le società di intelligenza artificiale stanno adottando misure per contribuire a mitigare questo rischio; Google Gemini, ad esempio, ha adottato limitazioni che impediscono agli utenti di chiedere informazioni sulle prossime elezioni in qualsiasi paese. Mentre l'intelligenza artificiale continua a evolversi, è necessario adottare misure per affrontare i potenziali rischi che questa tecnologia comporta per l'integrità delle elezioni statunitensi e garantire la fiducia del pubblico nel processo democratico.



Tutti gli occhi puntati sulle normative relative all'AI

Dato il suo potenziale impatto economico, i governi di tutto il mondo stanno lavorando attivamente per regolamentare l'AI e favorirne un utilizzo sicuro. Ad oggi, ci sono state almeno 1600 iniziative in questo senso in 69 paesi e nell'UE; si va da normative a strategie nazionali, sovvenzioni, investimenti e altro.^{14,15}

In generale, queste attività hanno l'obiettivo di comprendere l'impatto di questa tecnologia, stimolare l'innovazione e modellarne lo sviluppo responsabile attraverso le policy. Le normative sull'intelligenza artificiale continueranno a svilupparsi ed evolversi rapidamente, ma alcune recenti modifiche legislative possono essere utili per le aziende che cercano di comprendere queste tendenze.

Stati Uniti

Negli Stati Uniti, l'attenzione si è concentrata sull'ordine esecutivo della Casa Bianca sullo sviluppo e l'uso sicuri, protetti e affidabili dell'intelligenza artificiale,¹⁶ che obbliga gli sviluppatori dei più grandi sistemi di AI a riferire i risultati dei test di sicurezza anche al Department of Commerce e rivelare quando vengono utilizzate nuove grandi risorse di calcolo per il training di modelli AI. Ha inoltre richiesto a nove agenzie federali di completare valutazioni sul rischio dell'impatto dell'intelligenza artificiale sulle infrastrutture critiche. La Casa Bianca si concentra anche sull'utilizzo dell'AI per l'innovazione: nell'ambito dell'EO, il governo degli Stati Uniti ha istituito il programma pilota National Artificial Intelligence Research Resource (NAIRR) per consentire ai ricercatori statunitensi di accedere a potenza computazionale, dati e altri strumenti per sviluppare modelli di intelligenza artificiale.¹⁷

Resta da vedere se il governo degli Stati Uniti adotterà norme più stringenti sull'intelligenza artificiale. Ad oggi, almeno 15 aziende leader nel settore dell'intelligenza artificiale e quasi 30 aziende sanitarie hanno aderito agli impegni volontari della Casa Bianca per tutelare l'uso dell'AI.¹⁸ Nel frattempo, la FTC ha vietato l'uso dell'intelligenza artificiale per impersonare un'agenzia governativa o un'azienda, e vi è l'intenzione di estendere e espandere la norma alla protezione di privati e agenzie.¹⁹ La Casa Bianca starebbe inoltre esplorando la possibilità di richiedere watermark per i contenuti generati dall'intelligenza artificiale



14. OECD, [Policies, data and analysis for trustworthy artificial intelligence](#), 12 marzo 2024.

15. Deloitte, [The AI regulations that aren't being talked about](#), 12 marzo 2024.

16. Casa Bianca, [Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence](#), 30 ottobre 2023.

17. NAIRR Pilot, [The National Artificial Intelligence Research Resource \(NAIRR\) Pilot](#), 12 marzo 2024.

18. Reuters, [Healthcare providers to join US plan to manage AI risks – White House](#), 14 dicembre 2023.

19. Ufficio del procuratore generale della Pennsylvania, [FTC Bans Use of A.I. to Impersonate Government Agencies and Businesses](#), 26 febbraio 2024.



Unione Europea

Il Parlamento europeo ha recentemente approvato l'AI Act, che sarà la prima legislazione completa al mondo su questo tema, con una rigorosa serie di leggi e linee guida per diversi tipi di applicazioni AI classificate in base al rischio in diversi settori. La legge, che dovrebbe entrare in vigore nel 2026, richiederà che gli strumenti di intelligenza artificiale generici come ChatGPT rispettino i requisiti di trasparenza, come ad esempio che il contenuto sia stato generato dall'AI, che i modelli di training siano stati progettati per impedire la generazione di contenuti illegali, e che le aziende forniscano riepiloghi dei materiali protetti da copyright utilizzati per il training.

Le normative si faranno più stringenti per le applicazioni AI "ad alto rischio", come quelle utilizzate nei prodotti di consumo, inclusi giocattoli, aviazione, dispositivi medici e veicoli, e per l'AI che ha un impatto su aree particolari, come infrastrutture critiche, occupazione, affari legali, immigrazione e altro. L'UE vieterà completamente le applicazioni di intelligenza artificiale ritenute più rischiose, come quelle che utilizzano informazioni biometriche sensibili, cercano di manipolare il comportamento umano, utilizzano il riconoscimento emotivo per le assunzioni e l'istruzione o raccolgono immagini facciali non mirate su Internet o dalle telecamere a circuito chiuso.²⁰

Molti paesi stanno dando priorità agli investimenti sull'intelligenza artificiale. Singapore, ad esempio, ha annunciato un piano di investimenti da 740 milioni di dollari nell'ambito della National AI Strategy 2.0.²¹ Questo piano favorirà l'innovazione e consentirà l'accesso ai chip avanzati necessari per implementare l'AI, garantendo al tempo stesso che le imprese siano pronte a trarre vantaggio da questa tecnologia rivoluzionaria in centri di eccellenza con sede a Singapore.

20. Parlamento Europeo, [EU AI Act: first regulation on artificial intelligence](#), 19 dicembre 2023.

21. CNBC, [Singapore's AI ambitions get a boost with \\$740 million investment plan](#), 19 febbraio 2024.

Previsioni sulle minacce dell'AI

Secondo il World Economic Global Risk Report, la disinformazione e gli attacchi informatici generati dall'AI rappresentano due dei 10 principali rischi nel 2024, rispettivamente in posizione n. 2 e n. 5 della classifica.²²

Poiché il campo dell'intelligenza artificiale continua a evolversi rapidamente, anche nell'area della generazione di video e immagini, questi rischi non potranno che aumentare, così come aumenterà la nostra capacità di usare l'AI per mitigarli. Guardando al resto del 2024 e oltre, queste sono le principali previsioni sui rischi e sulle minacce dell'AI che vediamo all'orizzonte.

1 Il dilemma dell'AI utilizzata da gruppi associati a stati-nazione: minacce più sofisticate e blocco di contenuti antigovernativi

I gruppi di minacce associati agli stati utilizzeranno l'AI per generare minacce più sofisticate e cercheranno allo stesso tempo di bloccare l'accesso ai contenuti antigovernativi.

L'uso di strumenti di intelligenza artificiale da parte di gruppi di minacce associati agli stati non è un fenomeno nuovo, ma la sua traiettoria prevede una crescita significativa sia in termini di portata che di sofisticatezza.

I report di Microsoft e OpenAI confermano questa preoccupazione, rivelando che i gruppi di autori di minacce supportati da nazioni come Russia, Cina, Corea del Nord e Iran hanno approfondito e sfruttato attivamente le funzionalità di ChatGPT. Questo si estende a vari casi d'uso, tra cui spear phishing, generazione e revisione di codice e traduzione.

22. World Economic Forum, [Global Risks Report 2024: The risks are growing — but so is our capacity to respond](#), 10 gennaio 2024.

23. ZDNet, [Cybercriminals are using Meta's Llama 2 AI](#), 21 febbraio 2024.

Sebbene un intervento mirato abbia fermato alcuni di questi attacchi, le imprese devono prepararsi alla persistenza di iniziative di intelligenza artificiale associate agli stati. Questo comprende l'implementazione di strumenti di intelligenza artificiale popolari, la creazione di LLM proprietari e l'emergere di varianti senza restrizioni ispirate a ChatGPT, come FraudGPT o WormGPT. Il panorama dipinge un quadro difficile, in cui gli aggressori associati agli stati continueranno a sfruttare l'intelligenza artificiale in nuovi modi per creare minacce informatiche complesse e sconosciute.

2 Dark chatbot e attacchi basati sull'AI: la piaga dell'AI per scopi malevoli crescerà

Gli attacchi basati sull'AI potrebbero aumentare durante il corso dell'anno, perché il dark web è un terreno fertile per chatbot dannosi come WormGPT e FraudGPT, che amplificano le attività criminali.

Questi strumenti insidiosi diventeranno determinanti nell'esecuzione di tattiche di ingegneria sociale avanzata, truffe di phishing e varie altre minacce. Il dark web è sempre più popolare tra i criminali informatici che vogliono usare in modo illecito ChatGPT e altri strumenti di GenAI per una vasta gamma di attacchi informatici. Sono state identificate più di 212 applicazioni LLM dannose, che rappresentano solo una frazione di quanto è realmente disponibile sul dark web, e si prevede che questo numero crescerà in modo costante.

Imitando gli sviluppatori che utilizzano l'intelligenza artificiale generativa per aumentare l'efficienza, gli autori delle minacce utilizzano questi strumenti per scoprire e sfruttare le vulnerabilità, creare attacchi di phishing convincenti, eseguire campagne di vishing e smishing e automatizzare gli attacchi con maggiore velocità, sofisticatezza e scalabilità. Ad esempio, il gruppo di autori di minacce Scattered Spider ha recentemente utilizzato LLaMa 2 LLM di Meta per sfruttare la funzionalità di Microsoft PowerShell, consentendo il download non autorizzato delle credenziali dell'utente.²³ La traiettoria di questi progressi indica che le minacce informatiche inizieranno ad evolversi più rapidamente che mai, assumendo nuove forme che saranno più difficili da riconoscere e da cui sarà più complesso difendersi con le misure di sicurezza tradizionali.

3 Combattere l'AI con l'AI: le roadmap e le spese per la sicurezza includeranno difese basate sull'intelligenza artificiale

Le imprese adotteranno sempre più tecnologie di intelligenza artificiale per combattere gli attacchi informatici generati dall'AI, con particolare attenzione all'utilizzo del deep learning e dei modelli AI/ML per rilevare malware e ransomware nascosti nel traffico cifrato. I metodi di rilevamento tradizionali continueranno a lottare con i nuovi attacchi zero-day e il ransomware polimorfico (che può modificare il proprio codice per eludere il rilevamento). Gli indicatori basati sull'AI saranno quindi cruciali per identificare potenziali minacce. Questi strumenti svolgeranno inoltre un ruolo fondamentale nell'identificare e bloccare rapidamente il phishing e altri attacchi di ingegneria sociale generati dall'AI.

Le imprese incorporeranno l'AI sempre di più nelle loro strategie di sicurezza informatica. L'intelligenza artificiale sarà uno strumento fondamentale per ottenere visibilità sul rischio informatico e creare strategie concrete e quantificabili per stabilire le priorità e rimediare alle vulnerabilità. Ottenere informazioni concrete dalla vasta gamma di segnali spesso irrilevanti è da tempo una delle principali sfide per i CISO, perché la correlazione tra rischi e minacce può richiedere mesi se si utilizzano decine di strumenti diversi. Nel 2024, le aziende guarderanno quindi con interesse agli strumenti di GenAI per mettere ordine al caos, contrastare i rischi informatici e ottenere una sicurezza più snella ed efficiente.

4 Inquinamento dei dati nelle supply chain AI: il rischio che i dati dell'AI siano spazzatura aumenterà

L'inquinamento dei dati diventerà una delle principali preoccupazioni quando aumenteranno gli attacchi alla supply chain. Le aziende di AI, i loro modelli di training e i fornitori a valle saranno sempre più presi di mira da utenti malintenzionati.

La Top 10 di OWASP per le applicazioni LLM mette in evidenza l'inquinamento dei dati di training e gli attacchi alla catena di approvvigionamento come rischi significativi in grado di compromettere la sicurezza, l'affidabilità e le prestazioni delle applicazioni AI. Allo stesso tempo, le vulnerabilità nelle catene di approvvigionamento delle applicazioni AI, che conta partner di soluzioni tecnologiche, dataset di terze parti e plug-in o API di strumenti AI, offrono numerose possibilità di sfruttamento.

Le aziende che usano gli strumenti di intelligenza artificiale saranno sotto maggiore osservazione, perché assumono che questi strumenti siano sicuri e producano risultati accurati. Sarà essenziale garantire la qualità, l'integrità e la scalabilità dei dataset di training, in particolare per la sicurezza dell'AI.





5 Sfruttarne al massimo il potenziale o regolarne l'utilizzo: le imprese metteranno sulla bilancia la produttività e la sicurezza derivanti dall'utilizzo dell'AI

Ormai, molte aziende hanno superato le prime fasi di adozione e integrazione degli strumenti di intelligenza artificiale e considerato attentamente le proprie policy di sicurezza. Tuttavia, questa situazione è ancora in fase di sviluppo per la maggior parte delle aziende, e le domande su quali strumenti AI consentiranno, quali bloccheranno e come proteggeranno i propri dati rimangono aperte.

Poiché il numero di strumenti di intelligenza artificiale continua ad aumentare rapidamente, le aziende dovranno prestare molta attenzione alle implicazioni per la sicurezza; come minimo, dovranno ottenere informazioni approfondite sull'utilizzo dell'AI da parte dei propri dipendenti e implementare controlli di accesso granulari per reparto, team e utente. Le organizzazioni potranno anche cercare di ottenere informazioni più granulari sulle stesse app di intelligenza artificiale, ad esempio applicando policy di prevenzione della perdita di dati nelle app per impedire la fuga di dati sensibili o disabilitando alcune azioni degli utenti, come le attività di copia/incolla.

6 Inganno e distorsione con l'intelligenza artificiale: i deepfake virali alimenteranno le interferenze nelle elezioni e le campagne di disinformazione

Le tecnologie emergenti, come i deepfake, aprono la strada a minacce molto serie, come l'interferenza elettorale e la diffusione di contenuti di disinformazione. L'intelligenza artificiale è già stata utilizzata in modo dannoso durante le scorse elezioni statunitensi, ad esempio per generare robocall che hanno impersonato i candidati al fine di scoraggiare l'affluenza alle urne. Questi casi, sebbene allarmanti, rappresentano solo la punta dell'iceberg della disinformazione che l'intelligenza artificiale può causare.

Inoltre, l'uso dell'AI per questi fini potrebbe non essere limitato ad aggressori interni. Anche altre entità associate agli stati potrebbero sfruttare queste tattiche per seminare confusione e minare la fiducia nel processo elettorale. In un caso degno di nota, gli aggressori hanno utilizzato deepfake generati dall'intelligenza artificiale per indurre un dipendente a trasferire 25 milioni di dollari; questo dimostra l'impatto di questa tecnologia. nel mondo reale. Allo stesso modo, sui social media sono diventate virali immagini deepfake di celebrità come Taylor Swift, richiamando l'attenzione sulla facilità con cui i contenuti manipolati possono diffondersi prima che vengano implementate verifiche.

Caso di studio: come usare in modo sicuro ChatGPT nell'azienda

Best practice per l'integrazione dell'intelligenza artificiale e le policy di sicurezza aziendale.

Ormai, le aziende sono state molto esposte agli strumenti di intelligenza artificiale. Tuttavia, poiché il numero di applicazioni AI continua a crescere vertiginosamente, e l'adozione continua a ritmo sostenuto, le aziende possono adottare alcune best practice per proteggere i propri dati, dipendenti e clienti. Nel complesso, le aziende devono adattare in modo proattivo e continuo l'utilizzo dell'intelligenza artificiale e le strategie di sicurezza per rimanere al passo con l'evoluzione dei rischi e sfruttare al tempo stesso il potenziale di questa tecnologia.



CASO DI STUDIO

5 passaggi per integrare e proteggere gli strumenti di intelligenza artificiale generativa

Le imprese che cercano di implementare in modo sicuro le applicazioni AI devono adottare un approccio misurato. In generale, per prima cosa è possibile bloccare tutte le applicazioni AI per eliminare il rischio di perdere i dati, quindi si può procedere a implementare misure per adottare applicazioni AI specifiche e verificate in modo ponderato, con rigorosi controlli di sicurezza e degli accessi per mantenere il controllo completo sui dati aziendali. Per semplicità, il seguente percorso si concentra sull'LLM ChatGPT di OpenAI.

Fase 1: Bloccare tutti i domini e le applicazioni AI e ML

Per eliminare i rischi noti e sconosciuti associati alle migliaia di applicazioni AI disponibili, le organizzazioni possono adottare un approccio zero trust proattivo bloccando tutti i domini e le applicazioni AI ed ML nell'azienda. In questo modo, sarà possibile concentrarsi sull'adozione di poche applicazioni AI e controllarne attentamente i rischi.

Fase 2: Controllare e approvare selettivamente le applicazioni di intelligenza artificiale generativa

Successivamente, l'organizzazione dovrà identificare una serie di applicazioni di GenAI in grado di soddisfare standard elevati per determinati criteri, come la capacità di creare solide misure di protezione dei dati, di sicurezza e contrattuali al fine di proteggere i dati aziendali e dei clienti, oltre ad avere un grande potenziale trasformativo. Per molte aziende, ChatGPT sarà una di queste applicazioni.

Fase 3: Creare un'istanza del server ChatGPT privata nell'ambiente aziendale/DC

Per garantire il controllo completo sui propri dati, le organizzazioni dovrebbero ospitare ChatGPT in un tenant dedicato e sicuro (come un server AI privato di Microsoft Azure) all'interno dell'organizzazione. Quindi, attraverso controlli di sicurezza e obblighi contrattuali, le organizzazioni dovranno assicurarsi che né Microsoft né OpenAI (in questo esempio) abbiano accesso ai dati aziendali o dei clienti e che le query degli utenti non vengano utilizzate per addestrare ChatGPT su larga scala. In questo modo, l'organizzazione manterrà il controllo

sui propri dati di training, così gli utenti potranno ottenere risposte altamente pertinenti e accurate e si ridurrà al minimo il rischio di inquinamento dei dati da un data lake pubblico.

Passaggio 4: **Sposta l'LLM dietro il Single Sign-On (SSO) con una solida autenticazione a più fattori (MFA)**

Successivamente, l'organizzazione dovrebbe spostare ChatGPT dietro un'architettura proxy cloud zero trust, come Zscaler Zero Trust Exchange, per applicare controlli di sicurezza zero trust sull'accesso. Questo potrebbe includere anche lo spostamento di ChatGPT dietro un provider di identità (IdP) con una solida autenticazione SSO ed MFA che includa l'autenticazione biometrica. In questo modo, sarà possibile ottenere un accesso utente sicuro e veloce a ChatGPT e consentire all'azienda di configurare controlli di accesso granulari a livello di utente, team e dipartimento. Questa pratica garantirebbe inoltre un isolamento dei problemi a livello di utente, team e reparto.

Posizionare ChatGPT dietro un proxy cloud come Zero Trust Exchange consente all'organizzazione di ispezionare tutto il traffico TLS/SSL tra gli utenti e ChatGPT, in modo da rilevare minacce informatiche e fughe di dati applicano al tempo stesso sette livelli distinti di sicurezza zero trust.

Passaggio 5: **Applica il motore DLP Zscaler per prevenire fughe di dati**

Infine, l'organizzazione dovrebbe implementare un motore DLP per l'istanza ChatGPT, al fine di prevenire la fuga accidentale di informazioni critiche, come dati e codici proprietari, dati dei clienti, dati personali, dati finanziari/legali e altro. Questo garantisce che i dati altamente sensibili non lascino mai l'ambiente di produzione.

Seguendo questo percorso, gli utenti delle aziende possono sfruttare tutti i vantaggi di uno strumento di intelligenza artificiale generativa come ChatGPT, eliminando al tempo stesso i rischi più critici legati all'adozione di una tale applicazione.

Best practice per l'utilizzo dell'intelligenza artificiale

In generale, le aziende possono adottare alcune best practice per integrare gli strumenti di intelligenza artificiale nel business.

- **Valutare e mitigare continuamente i rischi degli strumenti AI** per proteggere la proprietà intellettuale, i dati personali e le informazioni dei clienti.
- **Garantire che l'uso degli strumenti di AI sia conforme alle leggi** e agli standard etici pertinenti, come le norme sulla protezione dei dati e le leggi sulla privacy.
- **Stabilire una chiara responsabilità per lo sviluppo e l'implementazione degli strumenti di intelligenza artificiale**, compresi i ruoli e le responsabilità definiti per la supervisione dei progetti AI.
- **Mantenere la trasparenza quando si utilizzano strumenti di intelligenza artificiale**, ovvero giustificare l'utilizzo e comunicarne chiaramente lo scopo alle parti interessate.

Linee guida alle policy sull'AI

Le imprese dovrebbero seguire queste best practice e stabilire un quadro di policy chiaro che definisca l'uso accettabile dell'AI a livello aziendale, l'integrazione e lo sviluppo dei prodotti, le policy sulla sicurezza e sui dati e le best practice dei dipendenti nell'utilizzo di strumenti di intelligenza artificiale. Le seguenti strategie possono costituire un utile punto di partenza per stabilire policy chiare sull'AI.

- **Non fornire ai modelli di intelligenza artificiale informazioni di identificazione personale (PII)** o informazioni non pubbliche, proprietarie o riservate.
- **L'intelligenza artificiale non può sostituire un essere umano** e non dovrebbe essere utilizzata per prendere decisioni senza un adeguato intervento umano.
- **I contenuti generati dall'intelligenza artificiale non dovrebbero essere utilizzati senza la revisione e l'approvazione umana**, soprattutto quando il contenuto è rappresentativo della tua organizzazione.
- **Lo sviluppo e l'integrazione di strumenti di intelligenza artificiale dovrebbero seguire un framework sicuro del ciclo di vita del prodotto (SPLC)** per garantire il massimo livello di sicurezza.
- **Esegui un'accurata due diligence dei prodotti prima di implementare soluzioni di intelligenza artificiale** e assicurati di valutarne le implicazioni etiche e di sicurezza.

In che modo Zscaler fornisce l'intelligenza artificiale e lo zero trust e mette in sicurezza l'AI generativa

Il potere trasformativo dell'intelligenza artificiale nella sicurezza informatica risiede nella sua capacità di essere utilizzata per combattere il panorama in evoluzione delle minacce AI. Zscaler usa l'intelligenza artificiale per aiutare le aziende a fermare gli attacchi in tutte le sue fasi e diagnosticare e mitigare facilmente i rischi.

La chiave per la sicurezza informatica basata sull'AI: dati di alta qualità su larga scala

Le aziende generano numerosissimi dati di log che possono contenere segnali altamente attendibili, i quali possono indicare potenziali percorsi per una violazione. Tuttavia, la difficoltà con cui si separano i segnali irrilevanti da quelli rilevanti ha storicamente reso difficile isolarli in modo rapido. Grazie all'intelligenza artificiale generativa, Zscaler può sfruttare questi dati per migliorare in modo efficace le misure di diagnostica e protezione ottenendo una comprensione delle vulnerabilità e dei punti deboli utilizzabili dagli aggressori. Questo non solo consente a Zscaler di prevedere le violazioni prima che si verifichino, ma offre anche ai dirigenti la possibilità di visualizzare e quantificare la maturità informatica e il rischio in modo olistico, dando priorità alle fasi di correzione con Zscaler Risk360.

Le funzionalità della GenAI non si estendono solo alla meta-analisi del rischio informatico aziendale, ma vengono anche inserite direttamente nei prodotti di sicurezza per rilevare e interrompere meglio le minacce avanzate lungo la catena di attacco. Direttamente integrati nel cloud di sicurezza più grande del mondo, i modelli LLM ed AI di Zscaler usano un data lake che registra oltre 390 miliardi di transazioni giornaliere, con oltre 9 milioni di minacce bloccate e 300 bilioni di segnali. Zscaler elabora quindi dati su larga scala altamente affidabili e intelligence sulle minacce per ottenere una sicurezza informatica AI iperconsapevole e accurata. Tutto questo porta a risultati di sicurezza informatica più efficaci per i professionisti IT e della sicurezza.





Sfruttare l'intelligenza artificiale lungo tutta la catena di attacco

Abbiamo discusso dei numerosi modi in cui gli autori delle minacce utilizzano l'intelligenza artificiale per lanciare minacce sofisticate a maggiore velocità e su vasta scala. Zscaler implementa funzionalità AI sulla piattaforma Zero Trust Exchange e nella suite di prodotti informatici per identificare e fermare sia gli attacchi basati sull'AI che quelli convenzionali in ogni loro fase.

Fase 1: Rilevamento della superficie di attacco

La prima fase di un attacco informatico coinvolge in genere gli autori delle minacce, che sondano la superficie di attacco aziendale connessa a Internet per identificare i punti deboli da sfruttare a proprio vantaggio. Spesso, questi includono vulnerabilità della VPN o del firewall, configurazioni errate o server senza patch. Rispetto al passato, l'intelligenza artificiale generativa ha reso questo compito molto più semplice per gli aggressori, che possono semplicemente interrogare un elenco di vulnerabilità note associate a queste risorse.

Sfruttando le informazioni ottenute con l'intelligenza artificiale in Zscaler Risk360, le aziende possono vedere immediatamente le applicazioni e risorse rilevabili (e quindi rischiose) sulla loro superficie di attacco connessa a Internet e nasconderle dietro Zero Trust Exchange, in modo che non siano visibili sulla rete Internet pubblica. In questo modo, si riduce istantaneamente e drasticamente la superficie di attacco aziendale, ed è possibile impedire agli aggressori di scoprire eventuali punti di ingresso.

Fase 2: Rischio di compromissione

Durante la fase di compromissione, gli aggressori cercano di sfruttare le vulnerabilità per ottenere l'accesso non autorizzato ai sistemi o alle applicazioni aziendali. Le innovazioni dell'intelligenza artificiale di Zscaler aiutano a ridurre il rischio di compromissione smantellando gli attacchi sofisticati e dando priorità alla produttività.

PREVENZIONE DEL PHISHING CON L'AI

I modelli di intelligenza artificiale di Zscaler rilevano siti di phishing noti e da paziente zero per prevenire il furto di credenziali e lo sfruttamento del browser; inoltre, analizzano modelli di traffico, comportamenti e malware per individuare infrastrutture di comando e controllo (C2) sconosciute in tempo reale. Questi modelli si basano sulla combinazione di intelligenze sulle minacce, ricerca di ThreatLabz e isolamento dinamico del browser. In questo modo, le organizzazioni possono rilevare in modo ancora più efficace nuovi attacchi di phishing, tra cui gli attacchi generati dall'AI e i domini C2.

DIFESA SANDBOX BASATA SU FILE CON L'AI

La sandbox Zscaler inline basata sull'intelligenza artificiale rileva istantaneamente i file dannosi mantenendo la produttività dei dipendenti. Le tradizionali tecnologie sandbox fanno attendere gli utenti mentre i file vengono analizzati, oppure assumono che vi sia un rischio da paziente zero se i file vengono consentiti al primo passaggio. La nostra tecnologia AI Instant Verdict identifica, mette in quarantena e previene istantaneamente i file dannosi con un elevato livello di sicurezza, anche per le minacce O-day, eliminando così la necessità di attendere che i file vengano analizzati. Questo include le minacce trasmesse tramite canali cifrati (TLS e HTTP) e altri protocolli di trasferimento file. I file innocui vengono invece consegnati in modo sicuro e immediato.

L'INTELLIGENZA ARTIFICIALE PER BLOCCARE LE MINACCE WEB

Zscaler Browser Isolation, basato sull'intelligenza artificiale, blocca le minacce zero-day garantendo al contempo ai dipendenti l'accesso ai siti corretti per lo svolgimento del proprio lavoro. Il filtraggio degli URL aziendali spesso richiede controlli più granulari rispetto ad autorizzazione/blocco; i siti bloccati spesso sono sicuri e necessari per il lavoro, e di conseguenza il team di assistenza si ritrova a gestire una serie di ticket inutili. Il nostro strumento AI Smart Isolation è in grado di identificare quando un sito può essere rischioso e di aprirlo in isolamento, trasmettendolo sotto forma di pixel in un ambiente sicuro e isolato. In questo modo, è possibile bloccare efficacemente le minacce web come malware, ransomware, phishing e download drive-by attraverso la creazione di un solido strato di sicurezza, senza richiedere alle aziende di bloccare automaticamente i siti.



Fase 3: Movimento laterale

Se gli aggressori ottengono l'accesso all'interno di un'organizzazione, proveranno a spostarsi lateralmente per accedere a dati e applicazioni sensibili. In molte organizzazioni, agli utenti è concesso di accedere a decine di applicazioni critiche; questo significa che la superficie di attacco interna è notevolmente estesa.

Le funzionalità AI di Zscaler riducono il potenziale raggio d'azione degli attacchi, analizzando i modelli di accesso degli utenti e suggerendo policy di segmentazione intelligente delle app per limitare il movimento laterale. Ad esempio, è comune vedere che, sul totale di coloro che hanno accesso a un'app, sono solo pochi quelli che ne hanno effettivamente bisogno. Zscaler può creare automaticamente un segmento di app che limita l'accesso solo a quei pochi dipendenti, riducendo così il rischio di movimento laterale di oltre il 99%.

Fase 4: Esfiltrazione dei dati

Nella fase finale di un attacco, gli aggressori lavorano per esfiltrare i dati sensibili. Zscaler utilizza l'AI per consentire alle organizzazioni di implementare la protezione dei dati più rapidamente. Il rilevamento dei dati basato sull'intelligenza artificiale elimina le lunghe attività di rilevamento e classificazione dei dati, che altrimenti potrebbero ritardare o impedire la distribuzione. L'AI di Zscaler rileva e classifica automaticamente tutti i dati di un'organizzazione, consentendo alle aziende di analizzare immediatamente le informazioni sensibili durante la configurazione di policy DPL (Data Loss Prevention) per impedire che i dati escano dall'organizzazione in caso di attacco o violazione.

Riepilogo delle offerte AI di Zscaler

Nell'ambito di Zero Trust Exchange, Zscaler Internet Access™ fornisce una protezione basata sull'AI per utenti aziendali, dispositivi e applicazioni web e SaaS in tutte le sedi, offrendo funzionalità di:

- **Rilevamento di phishing e delle attività di comando e controllo** contro siti di phishing e infrastrutture C2 sconosciute utilizzando il rilevamento inline basato sull'AI di Zscaler Secure Web Gateway (SWG).
- Sandboxing basato sull'intelligenza artificiale con prevenzione completa da malware e minacce O-day.
- **Policy dinamiche e basate sul rischio** con analisi continua del rischio relativo a utenti, dispositivi, applicazioni e contenuti per alimentare policy dinamiche di sicurezza e accesso.
- **Segmentazione basata sull'intelligenza artificiale** grazie a Zscaler Private Access™, con suggerimenti sulle policy di accesso automatizzate per ridurre al minimo la superficie di attacco e arrestare il movimento laterale utilizzando il contesto, il comportamento, la posizione dell'utente e la telemetria delle app private.
- Isolamento del browser basato sull'intelligenza artificiale, che crea una barriera sicura tra utenti e categorie web dannose, rendizzando il contenuto in un flusso di immagini per eliminare le fughe di dati e la diffusione di minacce attive.

INOLTRE, ZSCALER BLOCCA:

URL e IP osservati nel cloud Zscaler e fonti di intelligence open source e commerciali integrate in modo nativo. Sono incluse anche le categorie di URL ad alto rischio definite da policy che vengono comunemente utilizzate per il phishing, come i domini osservati e attivati di recente.

Firme IPS sviluppate dall'analisi di ThreatLabz condotta su kit e pagine di phishing.

Zscaler Risk360 offre un quadro di rischio completo e concreto che aiuta i leader aziendali e della sicurezza a quantificare e visualizzare il rischio informatico in tutta l'azienda.

La protezione dei dati con DLP e CASB offre la classificazione e la protezione dei dati basate sull'AI per tutti i canali, tra cui endpoint, email, workload, dispositivi personali e profilo cloud.

Advanced Threat Protection blocca tutti i domini C2 conosciuti.

Zscaler ITDR (Identity Threat Detection and Response) attenua il rischio di subire attacchi basati sull'identità, sfruttando la visibilità continua, il monitoraggio del rischio e il rilevamento delle minacce.

Zscaler Firewall estende la protezione C2 a tutte le porte e i protocolli, comprese le destinazioni C2 emergenti.

DNS Security difende dagli attacchi basati su DNS e dai tentativi di esfiltrazione.

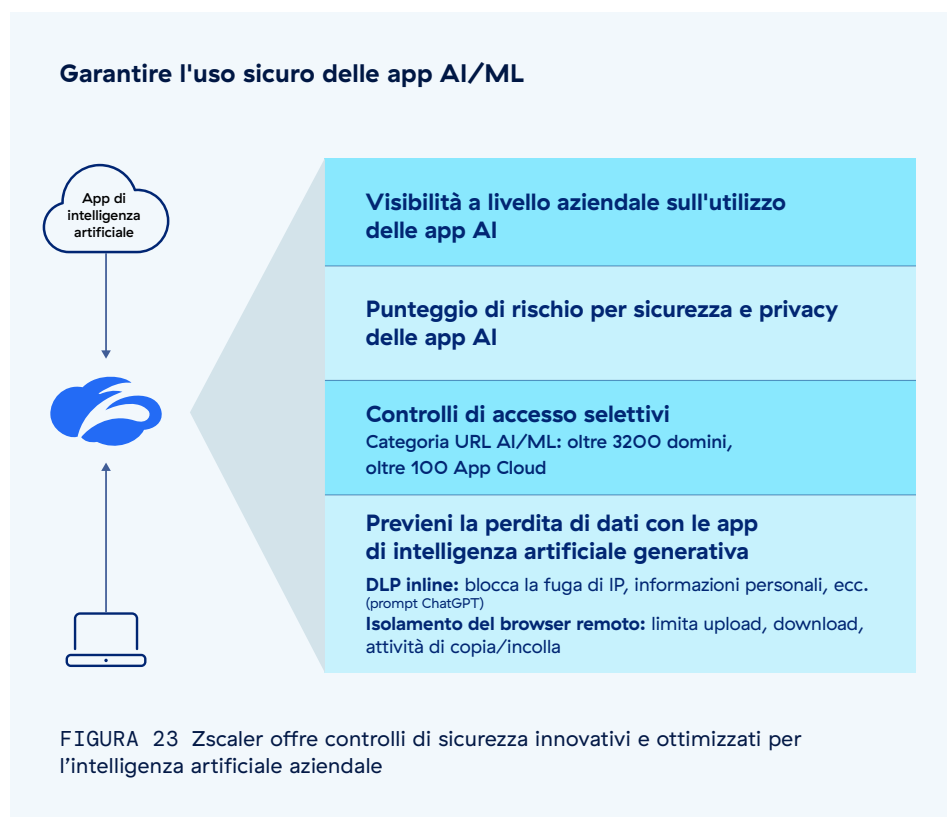
Zscaler Private Access™ protegge le applicazioni limitando il movimento laterale, sfruttando l'accesso a privilegi minimi, la segmentazione da utente ad app e l'ispezione completa inline del traffico delle app private.

AppProtection con Zscaler Private Access fornisce un'ispezione di sicurezza inline ad alte prestazioni dell'intero payload dell'applicazione per esporre le minacce.

Zscaler Deception™ rileva e blocca gli aggressori che tentano di spostarsi lateralmente o di incrementare i propri privilegi attirandoli con esche che sembrano server, applicazioni, directory e account utente.

Consentire la transizione verso l'intelligenza artificiale nelle aziende: il controllo è nelle tue mani

Zscaler offre alle imprese la possibilità di favorire l'innovazione, la creatività e la produttività con le applicazioni AI mantenendo al contempo utenti e dati al sicuro sui canali emergenti di esfiltrazione dei dati. Le imprese possono quindi [sfruttare il potenziale trasformativo dell'AI](#) per accelerare il proprio business senza bloccarne completamente le applicazioni e i domini.



ZSCALER CONSENTE ALLE AZIENDE DI:

- 01 Ottenere piena visibilità sull'utilizzo degli strumenti di intelligenza artificiale**
I dettagliati log forniscono una visibilità completa sul modo in cui i team utilizzano l'intelligenza artificiale; comprendono informazioni sulle applicazioni e i domini che visitano e dati e prompt utilizzati in strumenti come ChatGPT.
- 02 Creare policy flessibili per ottimizzare l'uso dell'intelligenza artificiale**
Il filtraggio URL efficace e personalizzato per le applicazioni AI ed ML consente di definire e applicare facilmente controlli di accesso e segmentazione granulari, bloccando l'accesso quando necessario e consentendolo con livelli di rischio accettabili grazie all'AI App Risk Scoring, il punteggio di rischio assegnato alle app AI. Le aziende possono consentire l'accesso a livello globale o in base a reparto, team o utente, oppure concederlo informando gli utenti dei rischi di questi strumenti. La segmentazione basata sull'AI semplifica l'identificazione dei segmenti di utenti appropriati per l'accesso a particolari applicazioni, riducendo al minimo la superficie di attacco interna associata a questi strumenti.
- 03 Applicare la sicurezza granulare dei dati per ChatGPT e altre applicazioni AI**
Le organizzazioni possono prevenire la fuga di dati sensibili caricati nelle applicazioni AI con i controlli granulari di Zscaler Cloud per l'intelligenza artificiale generativa. Applicando il motore DLP di Zscaler, le aziende possono garantire che nessun dato venga condiviso accidentalmente quando si utilizza uno strumento di AI. Inoltre, il rilevamento e la classificazione dei dati basati sull'AI consentono alle aziende di identificare e creare facilmente policy DLP relative ai dati più critici, inclusi il codebase aziendale, documenti finanziari e legali, dati personali, dati dei clienti e altro. [Questo video](#) mostra il modo in cui il motore di DLP impedisce agli utenti di inserire i dati della carta di credito su ChatGPT.
- 04 Implementare controlli efficaci utilizzando l'isolamento del browser**
Zscaler Browser Isolation rende le applicazioni AI in un ambiente sicuro, aggiungendo un livello di protezione che consente agli utenti di eseguire prompt e query agli strumenti AI limitando al contempo le attività di copia/incolla, gli upload e i download. In questo modo, è possibile mitigare il rischio che i dati sensibili vengano condivisi accidentalmente con strumenti di GenAI.

I leader delle imprese e della sicurezza sono a un bivio: devono implementare l'intelligenza artificiale e favorire l'innovazione per rimanere competitivi e al tempo stesso garantire che i dati non vengano violati. Zscaler consente alle aziende di affrontare questa transizione con sicurezza, grazie a una suite completa di controlli zero trust basati sull'AI che proteggono dagli attacchi, policy di intelligenza artificiale ottimizzate e strumenti di protezione dei dati che consentono di sfruttare tutto il potenziale dell'intelligenza artificiale generativa.

Appendice

Metodologia di ricerca di ThreatLabz

Il cloud di sicurezza globale Zscaler elabora oltre 300 bilioni di segnali giornalieri e blocca 9 miliardi di minacce e violazioni delle policy al giorno, con oltre 250.000 aggiornamenti di sicurezza giornalieri. Analisi di 18,09 miliardi di transazioni AI e ML da aprile 2023 a gennaio 2024 nel cloud Zscaler, Zero Trust Exchange.

Informazioni su Zscaler ThreatLabz

ThreatLabz è il team di ricerca sulla sicurezza di Zscaler. Questo team di esperti è responsabile della ricerca di nuove minacce e della protezione costante delle migliaia di aziende che utilizzano la piattaforma globale di Zscaler. Oltre alla ricerca sui malware e all'analisi del loro comportamento, i membri del team si occupano delle attività di ricerca e sviluppo di nuovi prototipi per la protezione contro le minacce avanzate sulla piattaforma Zscaler e conducono regolarmente controlli di sicurezza interni per garantire che i prodotti e l'infrastruttura di Zscaler siano in linea con gli standard di conformità. Sul suo portale, ThreatLabz pubblica regolarmente analisi approfondite sulle minacce nuove ed emergenti: research.zscaler.com.





Esplora il tuo mondo, in sicurezza.

Informazioni su Zscaler

Zscaler (NASDAQ: ZS) accelera la trasformazione digitale in modo che i clienti possano essere più agili, efficienti, resilienti e sicuri. Zscaler Zero Trust Exchange™ protegge migliaia di clienti dagli attacchi informatici e dalla perdita di dati, collegando in modo sicuro utenti, dispositivi e applicazioni in qualsiasi luogo. Distribuita in oltre 150 data center nel mondo, Zero Trust Exchange, basata sul framework SASE, è la piattaforma di cloud security inline più grande del mondo. Per saperne di più, visita il sito www.zscaler.it.

©2024 Zscaler, Inc. Tutti i diritti riservati. Zscaler™, Zero Trust Exchange™, Zscaler Internet Access™, ZIA™, Zscaler Private Access™ e ZPA™ e gli altri marchi commerciali indicati su [zscaler.it/legal/trademarks](https://www.zscaler.it/legal/trademarks) sono (i) marchi commerciali o marchi di servizio registrati o (ii) marchi commerciali o marchi di servizio di Zscaler, Inc. negli Stati Uniti e/o in altri Paesi. Tutti gli altri marchi commerciali sono di proprietà dei rispettivi titolari.