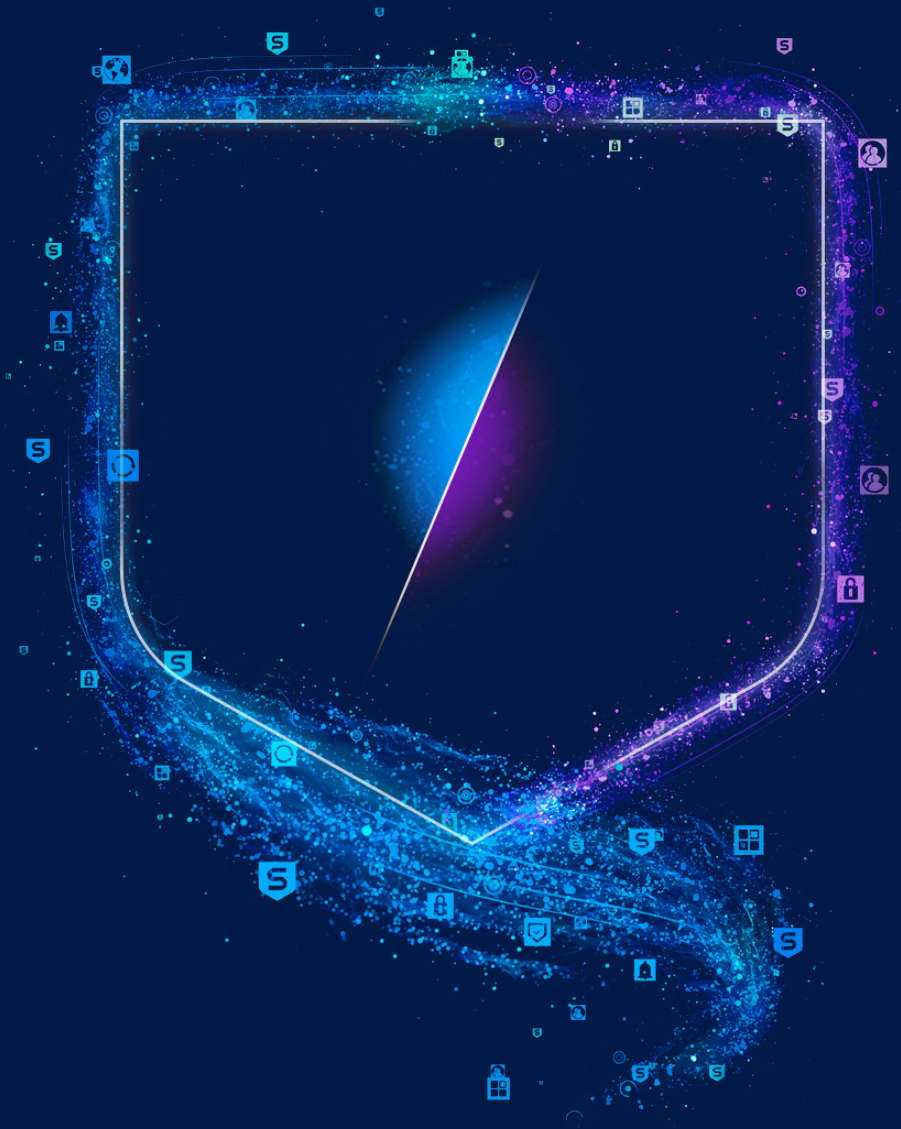
SOPHOS

Orientarsi nel clamore dell'IA nella Cybersecurity

Come sfruttare i vantaggi dell'IA in maniera sicura e consapevole, per potenziare le difese informatiche della tua azienda



Indice dei contenuti

Introduzione

Nel mondo della cybersecurity è ormai quasi impossibile evitare il clamore che circonda l'intelligenza artificiale. Le organizzazioni vengono costantemente bombardate con promesse invitanti di una trasformazione della sicurezza informatica grazie alle tecnologie di IA (maggiore protezione, costi più contenuti, necessità di meno personale) e gravi ammonimenti sul fatto che l'IA sta spianando la strada a una nuova era di attacchi informatici.

Questa guida è stata realizzata per aiutare le organizzazioni a orientarsi tra tutto il clamore e i pregiudizi sull'IA nella cybersecurity. Descrive quello che è (e quello che non è) in grado di fare l'IA per elevare le difese informatiche delle organizzazioni ed esplora i rischi operativi e di cybersecurity introdotti proprio dall'IA. La guida fornisce anche consigli su come mitigare questi rischi per sfruttare in maniera sicura e consapevole i vantaggi dell'IA, sia per potenziare la sicurezza informatica, sia per incrementare il ritorno sull'investimento.

Nelle pagine di questa guida troverai anche approfondimenti sulla realtà dell'utilizzo dell'IA, nonché sulle aspettative e sulle preoccupazioni che la circondano, in base ai risultati di un sondaggio vendor-agnostic condotto alla fine del 2024 coinvolgendo 400 IT e Cybersecurity Manager. Queste prospettive fornite da chi lavora in prima linea offrono un ottimo contesto e costituiscono il termine di paragone ideale per le organizzazioni che desiderano esplorare la propria posizione rispetto all'IA. Per i risultati completi, vedi Oltre al Clamore: la Realtà Aziendale dell'IA nella Cybersecurity

In ultima analisi, con o senza l'IA l'obiettivo rimane lo stesso: garantire il livello ottimale di resilienza informatica di cui ha bisogno la tua organizzazione per difendersi adeguatamente, riducendo allo stesso tempo la spesa totale. In altre parole, sfruttare il budget (immancabilmente limitato) di cybersecurity per offrire il giusto supporto all'azienda. Questa guida ti aiuterà a raggiungere questo obiettivo nell'era dell'intelligenza artificiale.

I vantaggi dell'IA nella cybersecurity

IA è un breve acronimo che definisce un'ampia gamma di opzioni in grado di supportare e accelerare la cybersecurity in vari modi diversi. La buona notizia è che ai team di sicurezza delle organizzazioni l'IA offre un vantaggio incrementale nettamente superiore rispetto a quello fornito agli autori degli attacchi. Due approcci molto comuni all'IA nel mondo della cybersecurity sono il deep learning e l'IA generativa.

Deep learning

I modelli di deep learning (DL) APPLICANO le nozioni apprese per svolgere attività. Sono in grado di accelerare l'applicazione di conoscenze con una rapidità che va ben oltre le capacità umane. Ad esempio, quando vengono addestrati adeguatamente, i modelli di DL riescono a identificare se un file è dannoso o innocuo in meno di un secondo, anche senza aver mai visto quel file prima.

Il DL è ideale per svolgere attività ripetitive su larga scala. Crea un modello *statistico* che considera i nuovi elementi nel contesto di tutto ciò che ha appreso dal suo vasto set di dati di apprendimento. I modelli di DL possono, ad esempio, analizzare milioni di campioni di file e determinare senza esitazione se contengono malware. Di conseguenza, il DL è ampiamente usato per elevare le capacità di protezione dei prodotti di cybersecurity.

I modelli di DL permettono ai team di sicurezza di gestire l'immensa quantità di minacce create dagli autori degli attacchi utilizzando l'automazione e il Cybercrime-as-a-Service. I modelli di DL possono anche essere aggiornati e adattati in base all'evoluzione degli attacchi, per tenere il passo con il panorama delle minacce.

Intelligenza artificiale generativa

I modelli di intelligenza artificiale generativa (GenAI) assimilano input e li utilizzano per CREARE nuovi contenuti. Alcuni esempi di applicazioni includono:

- Creare un riepilogo in linguaggio naturale dell'attività delle minacce fino a questo momento e fornire un elenco dei passaggi successivi consigliati agli analisti
- Portare alla luce informazioni dettagliate sul comportamento di un cybercriminale, analizzando i comandi che generano rilevamenti
- Permettere agli analisti di utilizzare ricerche in linguaggio naturale, invece di complicate query scritte in codice, per indagare sui rilevamenti sospetti
- Assegnare priorità alle patch da applicare, in base alla propensione di una vulnerabilità a essere soggetta a exploit

La GenAI è uno strumento potente per accelerare le Security Operations. Poiché svolge gran parte del lavoro richiesto per l'analisi dei dati, permette agli analisti di prendere rapidamente decisioni informate e di dedicare tempo ed energie agli ambiti nei quali possono avere maggiore impatto. In questo modo, spesso la GenAI è in grado di alleviare alcune delle pressioni che gravano sugli analisti, riducendo il rischio di burnout e di malcontento fra il personale. La GenAI può anche aiutare a ridurre i requisiti tecnologici delle Security Operations, permettendo agli analisti con meno esperienza di contribuire in maniera positiva e accelerare lo sviluppo delle proprie competenze.

Il viaggio verso la GenAI

La GenAI si basa sul trasformatore, una rete neurale di deep learning che analizza il contesto e la relazione tra i vari input (ad esempio le parole di una frase) e li usa per apprendere come creare output pertinenti. I trasformatori vengono utilizzati frequentemente in attività di elaborazione del linguaggio naturale (Natural Language Processing, NLP), come le traduzioni di testi o la formulazione di risposte alle domande. La "T" di ChatGPT sta infatti per "trasformatore".

Anche se l'uso di trasformatori è molto diffuso nella GenAI, non tutti i trasformatori sono generativi. Per esempio, BERT ("Bidirectional Encoder Representations from Transformers", ovvero Rappresentazioni Encoder Bidirezionali da Trasformatori) è un framework di machine learning open source per le NLP in grado di leggere il testo inserito in maniera bidirezionale, ovvero sia da destra verso sinistra che da sinistra verso destra. Questo approccio gli permette di migliorare significativamente la comprensione contestuale dei testi non classificati. Sophos usa BERT da diversi anni per identificare e neutralizzare gli attacchi Business Email Compromise.

Non esistono dimensioni universali

I modelli di IA hanno dimensioni molto variabili. I **Modelli di grandi dimensioni**, come Microsoft Copilot e Google Gemini, sono appunto modelli linguistici di grandi dimensioni (LLM, dall'inglese "Large Language Model") addestrati su un vasto set di dati per svolgere moltissime attività diverse. I modelli di piccole dimensioni vengono invece addestrati in base a un set di dati molto specifico e tipicamente sono progettati per svolgere un'unica attività, come rilevare URL o file eseguibili pericolosi. Poiché hanno un ambito di applicazione più limitato, i **modelli più piccoli** presentano vari vantaggi in termini di costi, velocità e performance, rispetto a quelli dei modelli più grandi.

Le limitazioni dell'intelligenza artificiale

Affidarsi esclusivamente all'intelligenza artificiale non è la risposta. O almeno non per il prossimo futuro. L'IA completa le capacità umane, ma non le rimpiazza interamente. Le minacce sono estremamente complesse e le Security Operations richiedono sia competenze tecniche che la capacità di applicare informazioni nel contesto dell'organizzazione. Da sola, l'IA non è in grado di anticipare le abili (e con ampie disponibilità finanziarie) organizzazioni cybercriminali del giorno d'oggi.

TIPO	**IA basata sul deep learning** *Applicazione* Utilizza reti neurali artificiali per riconoscere pattern e prendere decisioni con un processo che imita il funzionamento del cervello umano. APPLICA le informazioni apprese per svolgere attività. Esempio: **Rilevamento degli URL malevoli** Il modello di IA è addestrato per identificare i siti web pericolosi, permettendo così agli strumenti di sicurezza di bloccare l'accesso a questi siti	**Intelligenza artificiale generativa** *Creazione* Utilizza la struttura e i pattern dei dati esistenti per CREARE (generare) nuovi contenuti. Esempio: **Riepilogo di un caso di minaccia** Il modello di IA crea un riepilogo dell'attività della minaccia e fornisce agli analisti un elenco di azioni successive consigliate
DIMENSIONI	**Modelli di IA di grandi dimensioni** Strumenti multifunzione addestrati su ampie quantità di dati pubblicamente disponibili, e in grado di assistere con lo svolgimento di un ampio spettro di attività. Esempio: **Microsoft Copilot, Google Gemini**	**Modelli di IA di piccole dimensioni** Modelli con scopi precisi, che sono progettati, addestrati e compilati per casi di utilizzo specifici. Esempio: **Modello di rilevamento del malware Android**

Tasso di adozione dell'IA

L'IA è già ampiamente integrata nell'infrastruttura di cybersecurity di gran parte delle organizzazioni:

- Il 73% dichiara di usare soluzioni di cybersecurity che includono modelli di deep learning
- Il 65% sostiene di usare soluzioni di cybersecurity che includono funzionalità di IA generativa

Tuttavia, le applicazioni dell'IA nella cybersecurity non si limitano solo ai vendor esterni: il 34% delle organizzazioni afferma infatti di utilizzare già la GenAI internamente per potenziare le difese informatiche, ad esempio generando simulazioni di e-mail di phishing.

Con molta probabilità, entro breve l'adozione dell'IA sarà pressoché universale, visto che attualmente le organizzazioni che cercano una piattaforma di cybersecurity includono le funzionalità di IA nell'elenco dei requisiti nel 99% (cifra arrotondata) dei casi:

- Il 57% delle organizzazioni dichiara che le funzionalità di IA sono essenziali/estremamente importanti
- Il 41% sostiene che le funzionalità di IA sono importanti

Con questi tassi di adozione e utilizzo futuro, comprendere quali sono i rischi dell'IA nella cybersecurity e come mitigarli è una delle priorità principali per le organizzazioni di qualsiasi dimensione e ambito operativo.

73%
Usa strumenti di cybersecurity con modelli di deep learning

65%
Usa strumenti di cybersecurity con funzionalità GenAI

99%
Richiede funzionalità di IA quando deve scegliere una piattaforma di cybersecurity

La GenAI: grandi aspettative

Il clamore che circonda la GenAI ha fatto sorgere grandi aspettative su come questa tecnologia possa migliorare i risultati di cybersecurity. Il sondaggio ha rivelato qual è il principale vantaggio che le organizzazioni desidererebbero dalle funzionalità GenAI negli strumenti di cybersecurity, come indicato dalla seguente tabella.

Principale vantaggio desiderato dalle funzionalità di IA generativa Risultati con il maggior numero di risposte	
1=	Migliore protezione contro le minacce informatiche (20%)
1=	Maggiore ritorno sull'investimento nella cybersecurity (20%)
3	Maggiore impatto ed efficienza per gli analisti informatici (17%)
4	La certezza di tenere il passo con le innovazioni della cybersecurity (15%)
5=	Maggiore tranquillità nel sapere che la nostra organizzazione è adeguatamente protetta contro gli attacchi (14%)
5=	Minore rischio di burnout per i dipendenti, automatizzando le attività per concedere più tempo ai dipendenti che si occupano della cybersecurity (14%)

Quali eventuali vantaggi desidereresti dalle funzionalità di IA negli strumenti di cybersecurity? Risultati con il maggior numero di risposte (n=400)

Dall'ampio spettro di risposte ottenute emerge che non esiste un singolo vantaggio maggiormente desiderato dalla GenAI nella cybersecurity. Allo stesso tempo, le risposte più comuni riguardano il desiderio di una migliore protezione informatica o performance aziendale (sia in termini finanziari che operativi). I dati suggeriscono anche che l'inclusione di funzionalità GenAI nelle soluzioni di cybersecurity offrono maggiore tranquillità, nella certezza che l'organizzazione sta tenendo il passo con le più recenti funzionalità di protezione.

Il fatto che un minore rischio di burnout per i dipendenti si trovi al fondo di questa classifica sembra suggerire che le organizzazioni sono meno consapevoli o nutrono meno interesse nel potenziale della GenAI di offrire supporto agli utenti. Con l'attuale carenza di personale di cybersecurity, ridurre il malcontento tra i dipendenti è un ambito di focalizzazione importante, per il quale è possibile ricorrere all'aiuto dell'IA.

Una migliore **protezione** e un maggiore **ritorno sull'investimento** sono i principali vantaggi desiderati dalla GenAI per le organizzazioni

I rischi dell'IA nella cybersecurity

L'uso dell'IA nella Cybersecurity è una spada a doppio taglio. Se da un lato l'IA offre enormi vantaggi ai team di sicurezza nella lotta contro i cybercriminali, dall'altro introduce anche diversi rischi:

1. **Rischio in termini di minacce:** uso dell'IA negli attacchi informatici
2. **Rischio in termini di difese:** scarsa qualità dell'IA e implementazione inadeguata
3. **Rischio operativo:** eccessiva fiducia nell'IA
4. **Rischio finanziario:** scarso ritorno sull'investimento nelle tecnologie di IA
5. **Rischio di sabotaggio:** compromissione dei modelli pubblici di IA per mano dei criminali informatici

1. Rischio in termini di minacce: uso dell'IA negli attacchi informatici

Nonostante tutte le esagerazioni su come l'IA stia creando un nuovo panorama delle minacce, la realtà è meno drammatica. Le discussioni sull'IA nei forum dedicati al cybercrime sono limitate e molti autori di attacchi sono ancora scettici nei confronti dell'IA. I pochi tentativi di sviluppare malware, strumenti di attacco ed exploit utilizzando l'IA sono solitamente primitivi e di scarsa qualità.

Analogamente alle organizzazioni legittime, i cybercriminali sfruttano l'IA principalmente per migliorare la qualità dei contenuti e l'efficienza delle operazioni, anche se con obiettivi ben diversi. Per informazioni più dettagliate e aggiornate sul panorama delle minacce e sui più recenti attacchi basati sull'IA, visita il blog Sophos.

Migliorare la qualità dei contenuti

Una delle applicazioni più rapide, semplici e accessibili dell'IA negli attacchi informatici è elevare la qualità e la credibilità delle e-mail di phishing e di truffa per aumentare la probabilità che le vittime cadano in trappola durante gli attacchi.

"Indizi tipici", come errori di grammatica e di battitura o formati sospetti, possono essere eliminati facilmente con gli strumenti di IA. Con l'aiuto degli LLM pubblici, un'e-mail ben scritta per una campagna di phishing può essere creata in meno di un minuto. Analogamente, ora è semplicissimo creare testi e messaggi per i social media convincenti e scritti correttamente in qualsiasi lingua, per indurre i destinatari a cliccare su link malevoli o a condividere informazioni personali. Gli LLM aiutano gli autori degli attacchi anche a integrare facilmente informazioni opportune al momento giusto, aumentando ulteriormente la propensione delle vittime a cadere in trappola.

Gli strumenti di IA generativa hanno anche spianato la strada a una nuova era di truffe, nelle quali i cybercriminali fingono di essere un dirigente aziendale per indurre vittime ignare a effettuare bonifici bancari a loro favore. Le tecnologie di clonazione vocale sono ora talmente avanzate che, se addestrate correttamente, possono essere sfruttate dai criminali per convincere la vittima a credere di avere a che fare con la persona vera. In questi attacchi di phishing vocale, o "vishing", spesso un cybercriminale assume l'identità di un dirigente aziendale e chiama un dipendente per "chiedere" di effettuare operazioni illecite come acquistare carte regalo, fare un bonifico bancario o trasferire un file.

I malintenzionati sfruttano anche tecnologie deepfake basate sull'IA per imitare visivamente altre persone nel corso dei loro attacchi. I video deepfake vengono anche utilizzati per indurre dipendenti ignari a trasferire somme di denaro importanti, e per ingannare i programmi di riconoscimento facciale al fine di richiedere prestiti e aprire conti bancari.

Migliorare l'efficienza operativa

Molte aziende legittime utilizzano chatbot basati sull'intelligenza artificiale per migliorare l'esperienza dei loro utenti, e gli autori degli attacchi non sono da meno. Anche alcuni cybercriminali sfruttano LLM per migliorare i forum che frequentano, creando chatbot e risposte automatiche. In un esempio condiviso da Sophos X-Ops, il forum XSS aveva appositamente creato un chatbot per il sito che rispondeva alle domande degli utenti. L'amministratore scriveva (testo tradotto dal russo):

"In questa sezione puoi chattare con l'IA (Intelligenza Artificiale). Fai una domanda - Il nostro chatbot IA ti risponderà... Questa sezione e il bot IA sono progettati per risolvere problemi tecnici semplici, per l'intrattenimento dei nostri utenti [e] per acquisire familiarità con le possibilità dell'IA".

Compilare e addestrare modelli personalizzati richiede competenze di IA elevate, che sono costose e difficili da trovare. Anche se alcune gang di cybercriminali hanno esperti di IA interni, tipicamente gli autori degli attacchi sfrutteranno LLM già esistenti, invece di realizzarne uno nuovo.

Le risorse dei cybercriminali

È importante fornire il giusto contesto per l'uso dell'IA da parte degli autori degli attacchi. L'IA è solamente uno dei molteplici strumenti disponibili nell'arsenale dei cybercriminali. Da diversi anni questi hacker utilizzano automazione e modelli di tipo Cybercrime-as-a-Service per incrementare la portata e la frequenza degli attacchi. Per molte organizzazioni, l'esposizione ai rischi dovuti a queste funzionalità ha un impatto superiore rispetto a quello dei rischi derivati dall'uso dell'IA.

2. Rischio in termini di difese: scarsa qualità dell'IA e implementazione inadeguata

Come abbiamo visto, i modelli di IA sono già ampiamente integrati nelle difese informatiche delle organizzazioni. Anche se indubbiamente motivata da buone intenzioni, l'implementazione inadeguata di modelli di IA di scarsa qualità può introdurre inavvertitamente una tipologia completamente unica di gravi rischi di sicurezza. La propensione dei modelli di IA a introdurre rischi dipende da vari fattori, tra i quali:

- **La qualità dei dati su cui vengono addestrati questi modelli**. Il famoso detto "dati sbagliati, risultati sbagliati" si dimostra particolarmente veritiero nel caso dell'IA. L'uso di dati di scarsa qualità per addestrare i modelli rischia di introdurre errori, mentre l'utilizzo di set di dati non bilanciati ha il potenziale di alterare i risultati per via della rappresentazione eccessiva o insufficiente di certe variabili. Più dati di addestramento di alta qualità vengono adoperati, migliore sarà l'output generato.
- **Le competenze tecniche dei team che creano i modelli**. Per creare modelli di IA efficaci per la cybersecurity, occorre una comprensione approfondita di due ambiti diversi ma complementari:
  - **Le minacce:** per identificare le azioni che deve intraprendere un modello di IA, occorre prima capire le modalità di azione del malware e degli autori degli attacchi.
  - **L'intelligenza artificiale:** una volta che si sa quali sono le azioni che si desidera vengano svolte dall'IA, bisogna individuare e realizzare il giusto modello per raggiungere questo obiettivo.

  Per creare modelli di IA efficaci e con un impatto tangibile sulla cybersecurity, è fondamentale che questi due ambiti di competenza operino in stretta collaborazione, attingendo l'uno dalle capacità dell'altro.
- **La qualità del processo di sviluppo e implementazione del prodotto**. A metà del 2024, l'implementazione di un aggiornamento dei contenuti difettoso in un prodotto di sicurezza ha causato l'interruzione immediata dei servizi per molte aziende in tutto il mondo. In assenza di adeguati processi di test, valutazione e implementazione, le funzionalità di IA possono ipoteticamente causare danni ancora più gravi, con l'aggiunta del rischio che il problema potrebbe non essere facile da identificare o da risolvere.

Un falso senso di (cyber)sicurezza

Le organizzazioni sono generalmente consce del rischio che le soluzioni di cybersecurity potrebbero includere capacità di IA sviluppate o implementate in maniera inadeguata. La grande maggioranza (89%) dei professionisti IT e di cybersecurity che hanno partecipato al sondaggio sostiene di nutrire preoccupazioni legate alla potenziale presenza di difetti nelle funzionalità GenAI degli strumenti di sicurezza, che potrebbero arrecare gravi danni alla loro organizzazione: il 43% afferma infatti che questa possibilità sia "estremamente preoccupante", mentre il 46% indica di trovarla "piuttosto preoccupante".

Di conseguenza, non sorprende affatto che, nel 99% (cifra arrotondata) dei casi, le organizzazioni che considerano le funzionalità GenAI nelle soluzioni di cybersecurity valutino la qualità dei processi e dei controlli di cybersecurity utilizzati nello sviluppo della GenAI:

- Il 73% dichiara di svolgere una valutazione completa della qualità dei processi e dei controlli di cybersecurity
- Il 27% sostiene di svolgere una valutazione parziale della qualità dei processi e dei controlli di cybersecurity

Sebbene l'elevata percentuale di intervistati che afferma di svolgere una valutazione completa potrebbe sembrare inizialmente molto incoraggiante, in realtà questa statistica suggerisce che molte organizzazioni presentano un grave punto cieco in questo ambito.

La valutazione dei processi e dei controlli utilizzati per sviluppare funzionalità GenAI richiede trasparenza da parte del vendor e un certo livello di conoscenza dell'IA da parte di chi effettua la valutazione. Purtroppo, entrambe queste condizioni sono poco comuni. È infatti raro che i vendor delle soluzioni rendano disponibili i propri processi completi di sviluppo e implementazione delle loro funzionalità GenAI, e i team informatici delle aziende spesso hanno conoscenze limitate in termini di best practice di sviluppo per l'IA. Per molte organizzazioni, questa percentuale indica semplicemente che "non sanno quanto non sanno".

3. Rischio operativo: eccessiva fiducia nell'IA

L'IA influisce su quasi tutti gli aspetti delle nostre vite quotidiane, dal trovare il migliore itinerario per andare al supermercato, fino ai programmi TV consigliati. La sua natura onnipresente rende fin troppo facile ricorrere automaticamente all'IA per qualsiasi cosa e dare per scontato che l'IA sappia svolgere certe attività meglio degli esseri umani. Fortunatamente, molte organizzazioni sono consce e giustamente preoccupate delle implicazioni di cybersecurity derivate da un'eccessiva fiducia nell'IA:

- L'84% nutre preoccupazioni sulle conseguenti pressioni di ridurre il personale esperto in materia di cybersecurity
- L'87% nutre preoccupazioni sulla conseguente mancanza di responsabilità per la cybersecurity

Essere a conoscenza di questi rischi è il primo passo per mitigarli. È fondamentale ricordare che l'IA è solo uno degli strumenti disponibili nell'arsenale di difese informatiche di un'organizzazione: sebbene sia una parte importante dello stack di sicurezza, non è sempre l'approccio giusto, e raramente è la soluzione completa. Ogni organizzazione è diversa e l'uso dell'IA deve essere pertinente al contesto delle sue esigenze e della struttura aziendale più ampia.

4. Rischio finanziario: scarso ritorno sull'investimento nelle tecnologie di IA

Le funzionalità GenAI di alta qualità nelle soluzioni di cybersecurity sono costose da sviluppare e da mantenere. Gli IT e Cybersecurity Manager sono consci dell'impatto di questi costi, infatti l'80% ritiene che la GenAI causerà un aumento significativo dei prezzi dei prodotti di cybersecurity.

Nonostante si aspettino un aumento dei prezzi, nella maggior parte dei casi le organizzazioni vedono la GenAI come un modo per ridurre il costo totale della cybersecurity, con l'87% degli intervistati che si ritengono convinti che i costi della GenAI negli strumenti di cybersecurity verranno compensati interamente dai risparmi che offriranno.

Allo stesso tempo, le organizzazioni sono consapevoli che quantificare questi costi è una sfida. Le spese correlate alla GenAI sono tipicamente incluse nel prezzo totale dei prodotti e dei servizi di cybersecurity, il che rende difficile stabilire quanto stanno investendo le organizzazioni nella GenAI per la cybersecurity. In linea con questa assenza di certezze, il 75% dei partecipanti al sondaggio sostiene che questi costi sono difficili da misurare (il 39% si dichiara pienamente d'accordo, il 36% parzialmente d'accordo).

Senza una reportistica efficace, le organizzazioni rischiano di non ottenere il ritorno desiderato sull'investimento nell'IA per la cybersecurity, o ancora peggio di dirigere verso l'IA investimenti che otterrebbero migliori risultati altrove.

5. Rischio di sabotaggio: compromissione dei modelli linguistici di grandi dimensioni (LLM)

I rischi di cybersecurity implicati dall'IA si estendono oltre gli strumenti e le applicazioni di sicurezza. La rapida espansione globale dell'uso di LLM pubblici ha spianato la strada a cybercriminali sofisticati e con ampie disponibilità finanziarie, in grado di compromettere persino i modelli stessi per raggiungere i propri obiettivi. Questa situazione potrebbe evolversi in vari modi, con scenari che includono:

- **Data poisoning**. Nel loro articolo del 2023 Poisoning Web-Scale Training Datasets is Practical (Il poisoning dei set di dati di addestramento su scala web è pratico), Carlini e altri hanno dimostrato che il data poisoning (ovvero la manipolazione dei dati con cui viene addestrato il modello al fine di influenzarne gli output) è un rischio tangibile in termini di minacce.
- **Backdoor aggiunte da enti governativi**. Molte nazioni hanno a disposizione le risorse necessarie per creare potenti LLM. Aggiungendo backdoor segrete ai modelli, per poi renderli disponibili gratuitamente per l'uso pubblico, gli enti governativi possono, all'occorrenza, manipolare gli LLM a loro vantaggio.
- **Spoofing degli LLM**. I cybercriminali possono compromettere LLM legittimi (ad esempio aggiungendo backdoor) e poi promuovere le modifiche come "miglioramenti". Per ingannare gli utenti e convincerli a usare il loro strumento compromesso, falsificano il nome di un provider attendibile, ad esempio omettendo una lettera o usando la cifra "0" al posto della lettera "O".

Per un'indagine approfondita sulla compromissione degli LLM, dai un'occhiata ai più recenti dati di ricerca del team Sophos AI.

Consigli pratici per orientarsi nel clamore dell'IA

Anche se l'IA comporta dei rischi, adottando un approccio ponderato le organizzazioni possono orientarsi meglio e sfruttare in maniera sicura i vantaggi dell'IA per potenziare le proprie difese informatiche. Molti di questi consigli possono essere utilizzati anche per implementare l'IA in altri ambiti.

Rischio in termini di minacce: migliora le difese informatiche per affrontare l'era dell'IA

Uno degli obiettivi principali deve essere migliorare la resilienza alle minacce che sfruttano l'IA. Visto che i cybercriminali stanno usando l'IA principalmente per ottimizzare la qualità e la credibilità delle loro e-mail di phishing e di truffa, concentrarsi su questi ambiti è la scelta più logica. Ecco alcuni suggerimenti:

- **Migliorare la protezione delle e-mail**. Cerca soluzioni che siano in grado di rilevare le e-mail di phishing e di truffa generate con l'IA, per impedire che raggiungano le caselle di posta dei tuoi utenti.
- **Applica protezione contro gli attacchi Business Email Compromise e contro gli attacchi di imitazione di utenti VIP**. Scegli soluzioni di e-mail security che includano protezione contro attacchi BEC e VIP, come la capacità di analizzare i toni e lo stile dei contenuti per rilevare i tentativi di truffa.
- **Diffida soprattutto dei social media**: spesso gli utenti abbassano la guardia quando scorrono i contenuti dei canali social, quindi sono molto più propensi a cadere vittima di una truffa.
- **Applica processi in grado di mitigare il rischio di attacco tramite clonazione vocale**, ad esempio procedure da seguire se si ricevono richieste inattese di pagamento o condivisione di dati. Alcune opzioni includono:
  - Chiamare l'autore della richiesta per verificarla
  - Implementare l'uso di passcode o passphrase

Rischio in termini di difese: valuta la qualità dell'IA utilizzata nei prodotti di cybersecurity

Sii consapevole dei rischi e dell'impatto di un'IA scadente, quando scegli come investire il tuo budget di sicurezza. Chiedi ai vendor informazioni sui loro:

- **Dati di apprendimento**. Quali sono la qualità, la quantità e l'origine dei dati con cui vengono addestrati i modelli? Input di qualità implicano anche output di qualità.
- **Team di sviluppo**. Scopri da quali figure professionali è composto il team che ha creato i modelli. Quale livello di competenze sull'IA possiedono? Quali conoscenze hanno in materia di minacce, comportamenti dei cybercriminali e Security Operations?
- **Processi di sviluppo e implementazione del prodotto**. Quali processi di sicurezza applica il vendor quando sviluppa e implementa funzionalità AI nelle sue soluzioni? Quali attività di verifica e controllo svolge?

Infine, fatti questa domanda: "Quanta fiducia nutro nel fatto che questa organizzazione sia in grado di offrire un'adeguata soluzione di IA e di applicare controlli di qualità e implementazione affidabili?".

Rischio operativo: considera l'IA prima di tutto da un punto di vista umano

Se subisci una violazione, l'IA non ne sentirà l'impatto, ma i tuoi utenti sì. E se dovesse succedere il peggio e i tuoi sistemi venissero compromessi, avrai bisogno di un team di esperti in grado di capire la situazione e risolvere il problema nel contesto della tua azienda.

- **Mantieni la giusta prospettiva**. L'IA è solo uno degli strumenti disponibili nell'arsenale dei team di sicurezza. Usala liberamente, ma metti in chiaro che, dopotutto, la responsabilità per la cybersecurity è una responsabilità umana.
- **Non sostituire le tue risorse, potenziale**. L'attuale carenza globale di esperti di cybersecurity è un fatto ampiamente risaputo. I gravi problemi di burnout dei dipendenti non fanno che complicare ulteriormente il problema. Invece di cercare di ridurre il numero di dipendenti utilizzando l'IA, concentrati prima su come l'IA può essere d'aiuto al tuo personale. Automatizzando molte delle attività più ripetitive e meno tecniche delle Security Operations e offrendo approfondimenti mirati, l'IA può:
  - Liberare più tempo da dedicare a mansioni più importanti e con maggiore impatto aziendale
  - Ridurre il sovraccarico di avvisi, limitando così anche lo stress psicologico
  - Accelerare lo sviluppo professionale degli analisti più esperti
  - Permettere agli analisti meno esperti di svolgere attività di Security Operations, creando così una pipeline per il ricambio di risorse

Rischio finanziario: applica lo stesso rigore aziendale alle decisioni sugli investimenti nell'IA

Questo è uno dei rischi più semplici da mitigare per le organizzazioni, in quanto prevede molti fattori su cui hanno controllo.

- **Imposta degli obiettivi**. Imposta obiettivi chiari, specifici e dettagliati su quali risultati desideri dall'IA.
  - Identifica quello di cui hai bisogno. Quali lacune ha la tua azienda? Dov'è che l'IA può essere utile?
  - Considera i guadagni in termini di risorse finanziarie, tempo e protezione.
- **Quantifica i vantaggi**. Comprendi quanta differenza farebbero gli investimenti nell'IA.
  - Se il tuo obiettivo è ridurre i costi complessivi o il costo totale di proprietà della cybersecurity, quantifica i risparmi che ne deriverebbero.
  - Se desideri che l'IA diminuisca il malcontento nei team IT e di cybersecurity, indica chiaramente l'impatto specifico che lo strumento di IA avrebbe sul team. Quali attività rimuoverebbe dalle code? Quante ore di lavoro regalerebbe al tuo team?
- **Assegna priorità agli investimenti**. L'IA può essere d'aiuto in molti modi, alcuni dei quali avranno maggiore impatto rispetto ad altri. Identifica le metriche di particolare interesse per la tua organizzazione (risparmi finanziari, impatto sul morale del personale, riduzione dell'esposizione ecc.) e classifica le varie opzioni in ordine di importanza.
- **Misura l'impatto**. Le decisioni sugli investimenti vengono fatte con le migliori delle intenzioni. Assicurati di comprendere il rapporto tra la performance effettiva e le aspettative iniziali. Stai vedendo i vantaggi che attendevi? Ci sono benefici inattesi? Ed esistono ambiti in cui non sono stati ottenuti i risultati che speravi? Usa queste informazioni per apportare le modifiche necessarie.

Chiediti se l'IA è la strategia migliore per aiutarti a raggiungere il tuo obiettivo, oppure se un altro approccio o un'altra tecnologia avrebbero maggiore impatto.

Rischio di sabotaggio: mantieniti consapevole del pericolo

Questo è il rischio più difficile da mitigare per le organizzazioni. La semplice consapevolezza di questo pericolo aiuta a diminuirne l'impatto. Detto questo, quando scegli LLM pubblici, cerca:

- **Modelli realizzati da provider attendibili e molto conosciuti**. Anche se questi modelli non sono immuni agli attacchi di data poisoning, è più probabile che eventuali problemi di output vengano resi noti e condivisi pubblicamente.
- **I nomi esatti dei provider**. Gli autori degli attacchi falsificano i nomi di provider attendibili per ingannare gli utenti e indurli a pensare che i loro modelli compromessi siano in realtà quelli legittimi.

Gli esperti di IA per la cybersecurity si stanno impegnando attivamente per cercare modi per neutralizzare questo rischio.

Conclusione

L'IA offre enormi vantaggi per la cybersecurity. Evitando tutto il clamore che circonda l'IA e adottando un approccio ponderato e basato sui risultati, le organizzazioni possono sfruttare questa tecnologia per potenziare le proprie difese informatiche e fornire ottime risorse ai loro preziosissimi esperti IT e di cybersecurity.

Informazioni sul sondaggio

Fonte: Oltre al Clamore: la Realtà Aziendale dell'IA nella Cybersecurity

Sophos ha affidato a Vanson Bourne, un'azienda indipendente specializzata nel campo della ricerca, il compito di condurre un sondaggio che ha coinvolto 400 IT e Cybersecurity Manager in organizzazioni con 50-3.000 dipendenti. Il sondaggio è stato svolto nel mese di novembre 2024 e gli intervistati provengono da 13 settori diversi. Per garantire un'ampia rappresentanza, il sondaggio è vendor-agnostic, con organizzazioni partecipanti che utilizzano soluzioni di protezione endpoint di 19 vendor diversi.

Informazioni su Sophos

Sophos è leader globale nella cybersecurity, con una gamma completa di prodotti e servizi di cybersecurity pluripremiati, che variano da firewall, protezione endpoint e strumenti di EDR/XDR, fino a servizi di Managed Detection and Response (MDR) e Incident Response (IR).

Sophos eleva la cybersecurity con l'intelligenza artificiale dal 2017, abbinando all'IA le competenze umane dei nostri esperti, per bloccare un'ampia gamma di minacce, indipendentemente da dove vengano eseguite. I nostri prodotti e servizi includono funzionalità di deep learning e IA generativa in grado di risolvere i problemi più critici per conto dei clienti, e le nostre soluzioni vengono fornite attraverso la più estesa piattaforma di sicurezza AI-native del settore. Addestrata con i dati provenienti dagli attacchi osservati in oltre 600.000 ambienti dei clienti, la nostra piattaforma di IA adattiva garantisce una protezione senza paragoni contro le minacce avanzate ed estende il potenziale dei team di sicurezza interni delle organizzazioni.

Per saperne di più e per esplorare le soluzioni Sophos, visita www.sophos.it

2025-01-15 (WP-MP)

SOPHOS