CyberGrant
SECURITY MADE EASY

WHITE PAPER

AI & Cybersecurity

Sommario

Introduzione

L’AI viene sempre più impiegata nella cybersecurity per migliorare la protezione di sistemi informatici, di reti e di dati dai cyber attack, automatizzando il rilevamento delle minacce, analizzando grandi volumi di dati, identificando modelli e rispondendo agli incidenti di sicurezza in tempo reale.

Tuttavia, può essere altresì utilizzata come arma potente dai cyber criminali per sferrare attacchi sempre più sofisticati.

Evoluzione delle minacce

Gli attacchi basati sull'IA possono aggirare le misure di sicurezza tradizionali, automatizzare le attività dannose e sfruttare vulnerabilità su scala. Vediamo di seguito di che si tratta.

Deepfake

I deepfake sono tra gli strumenti con il maggiore impatto visivo e psicologico a disposizione dei cyber criminali.

Di fatto, essi utilizzano modelli di apprendimento automatico avanzati come le reti avversarie generative (GAN - Generative Adversarial Networks), per **creare video e clip audio iperrealistici** in grado di convincere la maggior parte delle persone della loro autenticità.

Ecco alcuni delle modalità con cui i cyber criminali utilizzano i deepfake per attività illecite:

- **Spionaggio aziendale** - I deepfake possono essere utilizzati per **impersonare dirigenti aziendali** e decisori, autorizzando transazioni fraudolente, inducendo i dipendenti a **divulgare dati aziendali sensibili** o diffondendo false informazioni per manipolare i prezzi delle azioni.
- **Ricatto ed estorsione** - I criminali informatici possono fabbricare video dannosi/ compromettenti di un individuo e **minacciarlo di rilasciarlo se non viene pagato un riscatto**.
- **Diffusione di notizie false** - I deepfake possono essere utilizzati per **creare interviste o discorsi falsi** di personaggi importanti, quali politici o funzionari della sanità pubblica, diffondendo disinformazione e minando la fiducia nelle istituzioni.

È doveroso evidenziare che **l'individuazione dei deepfake presenta delle sfide più ardue**, poiché i metodi forensi tradizionali faticano a tenere il passo con il realismo dei deepfake moderni. Inoltre, gli strumenti di rilevamento devono analizzare microespressioni, incongruenze nell'illuminazione e discrepanze audio-visive, attività che richiedono a loro volta un'AI avanzata.

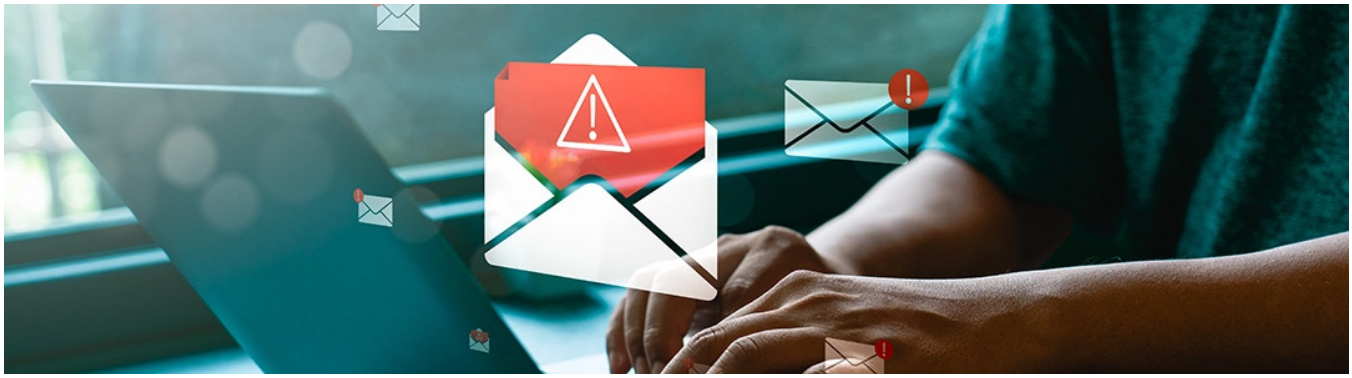
Phishing

> *"l'AI ha reso le truffe di phishing molto più efficaci"*

Il phishing si è evoluto da semplici **truffe via e-mail ad attacchi altamente sofisticati,** potenziati dall'IA e che sono difficili da identificare e di cui è facile cadere vittime. Di fatto, l'AI ha reso le truffe di phishing molto più efficaci per le seguenti caratteristiche:

- **Competenze linguistiche - I Large Language Model (LLM) – tra cui ChatGPT - possono generare messaggi grammaticamente perfetti** e contestualmente pertinenti e che si adattano a diversi stili di comunicazione aziendale.
- **Generazione di contenuti dinamici - L'IA può personalizzare le e-mail di phishing in tempo reale** in base al comportamento online del destinatario, al ruolo lavorativo o alle attività recenti.
- **Minaccia multilingue – L'IA può tradurre i contenuti di phishing in più lingue, mantenendo le sfumature culturali** e, al contempo, aumentare il bacino dei bersagli dei cyber criminali a livello internazionale.

Inoltre, si segnala tra gli esempi comuni di phishing, basate sull'IA, le email generate dall'IA che sembrano provenire dai dipartimenti delle Risorse Umane e che chiedono ai dipendenti di aggiornare le proprie credenziali, nonché fatture o richieste di pagamento false che sembrano identiche a quelle provenienti da



Malware e RaaS (Ransomware-as-a-service) basato sull'AI

Il malware basato sull'IA rappresenta un'evoluzione significativa: a differenza delle varianti tradizionali con modelli statici, **può adattarsi dinamicamente agli ambienti target** e analizzare le misure di sicurezza in tempo reale, **modificando le tattiche per aggirare le difese.**

Tali malware avanzati perfezionano continuamente le strategie durante l'esecuzione, diventando progressivamente più difficili da rilevare.

Parallelamente, il ransomware si è evoluto nel modello Ransomware-as-a-Service (RaaS), democratizzando l'accesso alle tecniche criminali e l'integrazione dell'AI ha amplificato questa minaccia, abilitando auto-targeting intelligente, aggiramento automatizzato della sicurezza, adattamento contestuale e apprendimento continuo, rendendo gli attacchi sempre più impattanti per le organizzazioni.

Social engineering

Il social engineering si basa sullo **sfruttamento delle emozioni umane e dei pregiudizi cognitivi. I cyber criminali,** grazie all'AI generativa, **possono automatizzare e personalizzare** queste manipolazioni su larga scala e, precisamente essi sono in grado di:

- **Creare fiducia** - L'AI può simulare conversazioni di lunga durata, creando gradualmente un rapporto di fiducia con un bersaglio, prima di mettere in atto una truffa.
- **Incutere paura e urgenza** - I messaggi generati dall'IA possono creare un senso di panico - i.e.: "Il tuo account è stato compromesso!" – spingendo i destinatari a prendere decisioni affrettate che vanno a loro discapito.
- **Fare leva sull'autorità** - Le comunicazioni generate dall'AI possono impersonare i superiori di un dipendente, spingendolo a eludere i protocolli di sicurezza.

È doveroso ricordare che i cyber criminali possono anche creare **falsi profili sui social media con foto e storie personali generate dall'AI,** oppure programmare bot di AI che partecipano a conversazioni sui social per raccogliere informazioni e influenzare le opinioni.

AI pubblica e rischio di esfiltrazione dati

> *"Le informazioni sensibili possono diventare accessibili ad altri utenti attraverso le risposte generate"*

L'adozione crescente di strumenti di AI pubblici – i.e. **chatbot, assistenti virtuali, ChatGPT,** ecc. -ha introdotto nuovi rischi significativi per la sicurezza dei dati aziendali. Quando i dipendenti utilizzano piattaforme di AI pubbliche per attività lavorative, possono **inconsapevolmente condividere informazioni sensibili o riservate**. Tali dati, una volta inseriti nei sistemi di AI, rischiano di: essere utilizzati per l'addestramento dei modelli; **diventare accessibili ad altri utenti attraverso le risposte generate**; essere conservati sui server dei fornitori, facendo perdere il controllo aziendale su informazioni critiche.

Le tipologie di dati a rischio includono: **codice sorgente, strategie di business, informazioni finanziarie, dati dei clienti e proprietà intellettuale**. Particolarmente preoccupante è la mancanza di consapevolezza degli utenti: molti dipendenti, attratti dalla praticità di questi strumenti, non comprendono di star esponendo asset aziendali critici. Inoltre, le politiche di privacy dei servizi di AI pubblica variano considerevolmente e non sempre garantiscono la riservatezza necessaria.

Come si muovono i player della cybersecurity

31,70%

Il CAGR previsto tra il 2025 e il 2032

Fortune Business Insight

Il mercato globale delle soluzioni di cybersecurity basate sull'AI - secondo quanto rivela Fortune Business Insight - è stato valutato a **26,55 miliardi di dollari nel 2024 e si prevede che crescerà da 34,10 miliardi di dollari nel 2025** a 234,64 miliardi di dollari entro il 2032, registrando un CAGR del 31,70% durante il periodo di previsione.

Le principali aziende del settore stanno investendo massicciamente nello sviluppo di piattaforme AI-driven per il rilevamento delle minacce, la risposta automatizzata agli incidenti e l'analisi predittiva, integrando machine learning, elaborazione del linguaggio naturale e algoritmi di deep learning per **identificare pattern anomali, automatizzare i processi di threat hunting e ridurre drasticamente i tempi di risposta agli attacchi**.

Tuttavia, l'adozione dell'AI nella cybersecurity presenta limiti significativi, dato che i modelli di machine learning richiedono **enormi quantità di dati per l'addestramento, con il rischio di produrre falsi positivi** che sovraccaricano i team di sicurezza.

Inoltre, la dipendenza da dataset storici può rendere i sistemi vulnerabili ad attacchi completamente nuovi, mentre gli stessi algoritmi AI possono essere manipolati attraverso tecniche di adversarial machine learning. Ancora, la "black box" di alcuni modelli rende difficile spiegare le decisioni automatizzate, complicando le attività di audit e di incident response.

Sul fronte della compliance, l'integrazione dell'AI solleva questioni critiche rispetto alle seguenti normative vigenti:

GDPR - Il regolamento impone requisiti stringenti sul trattamento dei dati personali, richiedendo trasparenza nelle decisioni automatizzate e garantendo il diritto alla spiegabilità.

NIS2 - La direttiva richiede misure di gestione del rischio proporzionate e verificabili.

DORA - Il regolamento per il settore finanziario impone test di resilienza operativa che devono includere la validazione dei sistemi AI.

A fronte di quanto sopra è evidente che le aziende devono essere in grado di bilanciare l'innovazione tecnologica con obblighi di data minimization, privacy-by-design e accountability, documentando accuratamente i processi decisionali dell'AI, oltre a garantire la supervisione umana nelle operazioni critiche. Pertanto, la sfida principale resta armonizzare l'automazione intelligente con i principi di trasparenza, responsabilità e controllo richiesti dal framework normativo europeo.

Ambiti di applicazione dell'AI in cybersecurity

È doveroso evidenziare che, se da un lato l'AI generativa diventa uno strumento a disposizione degli aggressori, dall'altro lato, i difensori sfruttano la stessa tecnologia per creare sistemi di sicurezza informatica più intelligenti, veloci e adattabili, offrendo vantaggi significativi per la cybersecurity, grazie all'automatizzazione del rilevamento delle minacce e delle risposte in tempo reale a potenziali attacchi.

Vediamo in dettaglio di che si tratta.

Catalogazione dei dati e prevenzione esfiltrazione e data breach

L'AI offre strumenti avanzati per la protezione dei dati aziendali attraverso la **classificazione automatica e intelligente delle informazioni**. I sistemi basati su AI possono analizzare e catalogare automaticamente i dati secondo il livello di sensibilità, assegnare diritti di accesso appropriati e implementare controlli dinamici che si adattano al contesto e al comportamento degli utenti. Tali sistemi rilevano **pattern anomali di accesso, identificano potenziali tentativi di esfiltrazione in tempo reale e applicano policy di sicurezza granulari**.

Inoltre, l'apprendimento automatico consente di migliorare continuamente l'accuratezza nella rilevazione delle minacce, riducendo significativamente i rischi di data breach e garantendo conformità normativa attraverso un monitoraggio proattivo e automatizzato degli asset informativi.

Rilevamento e risoluzione delle minacce

L'AI permette di garantire:

- **Controllo degli standard del codice –** L'AI può analizzare il codice sorgente per verificarne la conformità alla sicurezza e rilevare potenziali vulnerabilità nelle prime fasi del ciclo di sviluppo.
- **Identificazione delle vulnerabilità -** Gli algoritmi di apprendimento automatico possono valutare i software e i sistemi alla ricerca di punti deboli e stabilire le priorità per la correzione.
- **Fornire soluzioni** - Gli strumenti basati sull'AI possono suggerire soluzioni o applicare automaticamente patch ai sistemi vulnerabili.

Rilevamento di attacchi di hacking e malware

I sistemi di AI possono monitorare il **traffico di rete e analizzare i modelli comportamentali per rilevare tentativi di hacking e infiltrazioni di malware**. L'AI, sfruttando l'apprendimento automatico, può identificare anomalie e fornire avvisi in tempo reale, consentendo un intervento rapido per mitigare potenziali minacce.

Monitoraggio del comportamento e rilevamento delle anomalie

L'AI eccelle nel monitoraggio del comportamento attraverso l'analisi dei normali modelli di comportamento del sistema e degli utenti.

Rilevamento e prevenzione delle intrusioni

Si tratta di implementare soluzioni di cybersecurity basate sull'AI, tra cui:

- **Network-based Intrusion Detection and Prevention (NIDP)** – Si tratta di una soluzione che monitora il traffico di rete per rilevare attività dannose.
- **Host-based Intrusion Detection and Prevention (HIDP)** - Si tratta di una soluzione che analizza l'attività sui singoli dispositivi per identificare potenziali minacce.
- **Sistemi di rilevamento e prevenzione delle intrusioni (IDPS)**- Si tratta di sistemi che combinano strategie basate su rete e host per fornire una copertura di sicurezza completa.

Inoltre, l'AI migliora il rilevamento e la prevenzione delle intrusioni in termini di:

- **Monitoraggio in tempo reale**, i.e.: sorveglianza continua di sistemi e reti.
- **Rilevamento delle anomalie**, i.e.: identificazione delle deviazioni dai modelli di comportamento stabiliti.
- **Risposta automatica**, i.e.: adozione di misure immediate e automatizzate per contenere le minacce.
- **Analisi predittiva**, i.e.: previsione e prioritizzazione delle potenziali minacce sulla base di tutti i dati storici e delle tendenze contingenti.

Analisi in tempo reale e risposta agli incidenti

> *"l'AI è in grado di analizzare in tempo reale degli eventi di sicurezza e automatizzare i processi di triage degli incidenti"*

Gli strumenti basati sull'AI forniscono **analisi in tempo reale degli eventi di sicurezza e automatizzano i processi di triage degli incidenti.** Inoltre, essi svolgono un ruolo fondamentale nella risposta agli incidenti attraverso:

- **Rilevamento precoce** – L'AI contribuisce a identificare le minacce prima che causino danni significativi.
- **Risposta rapida** – L'AI garantisce la mitigazione automatica dei rischi tramite protocolli predefiniti.
- **Indagine automatizzata** – L'AI è in grado di utilizzare le informazioni sulle minacce per determinare la portata e l'impatto di un attacco e condividere report approfonditi e personalizzati, nonché fornire tutte le info utili all'analisi forensics.
- **Analisi comportamentale** – L'AI è in grado di valutare il comportamento dell'utente e del sistema per identificare potenziali rischi.

Identificazione della fonte e della causa degli incidenti di sicurezza

L'AI, subito dopo un attacco alla sicurezza, l'AI offre il suo prezioso contributo per **mitigare i rischi in una certa misura**, dato che è in grado di garantire di versi aspetti, come qui di seguito riportati.

Protezione del sistema interessato

- **Isolamento dei sistemi** - I sistemi basati sull'AI possono rilevare automaticamente i dispositivi compromessi e isolarli dalla rete per impedire lo spostamento laterale delle minacce.
- **Contenimento automatizzato delle minacce** – L'utilizzo di strumenti basati sull'AI, quali l'EDR (Endpoint Detection and Response), consente di bloccare automaticamente le attività sospette.
- **Controllo dinamico degli accessi** - I modelli di AI possono applicare misure di autenticazione adattive e bloccare gli account interessati.

Documentazione dell'incidente

- **Registrazione automatica** - L'AI può acquisire e organizzare i registri da varie fonti in tempo reale, garantendo che nessuna informazione critica venga persa.
- **Riconoscimento di modelli** - Gli algoritmi di apprendimento automatico possono identificare attività anomale nei registri, semplificando il processo di documentazione.

Preservare le prove

- **Controlli dell'integrità dei dati** - Gli strumenti di AI possono applicare hash crittografici ai file per verificarne l'integrità.
- **Backup automatici** - I sistemi di AI, durante un incidente, possono

creare snapshot di sistemi e file per l'analisi forense.

- **Gestione della catena di custodia** - L'AI può aiutare a mantenere l'autenticità e la tracciabilità delle prove digitali.

Esecuzione dell'analisi iniziale dell'incidente

- **Triage degli incident** - I sistemi di AI possono stabilire la priorità degli incidenti in base alla gravità, all'impatto e all'urgenza.
- **Integrazione dell'intelligence sulle minacce** - L'AI può correlare l'incidente con database di minacce esterne per identificare modelli di attacco noti.

Recupero dei dati cancellati

- **Tecniche di ricostruzione dei dati** - L'AI può utilizzare modelli predittivi per ricostruire dati persi o danneggiati.
- **Strumenti di incisione dei file** - Gli algoritmi di apprendimento automatico possono riconoscere e ripristinare frammenti di file da immagini disco.

Analisi comportamentale e di rete del malware

- **Sandbox basati sull'AI per l'esecuzione controllata del malware** - I sandbox basati sull'AI eseguono malware in ambienti virtuali isolati, osservando le azioni del software dannoso in modo sicuro per comprenderne funzionalità e impatto. L'AI automatizza l'analisi e identifica modelli di comportamento dannoso.
- **Monitoraggio delle interazioni del malware con il sistema** - L'obiettivo è rilevare come il malware interagisce con il sistema, incluse modifiche ai file, variazioni del registro, connessioni di rete e tentativi di sfruttare vulnerabilità.
- **Confronto con minacce note e rilevamento delle anomalie** - I sistemi di AI monitorano il comportamento del malware confrontandolo con minacce note e utilizzando il rilevamento delle anomalie per identificare nuovi modelli di attacco.
- **Risultati dell'analisi AI** - È possibile creare firme malware, sviluppare contromisure per attacchi futuri e fornire informazioni per l'intelligence sulle minacce.
- **Identificazione di modelli di rete sospetti** - L'AI monitora e identifica modelli di rete sospetti che indicano una violazione della sicurezza.
- **Analisi sulla firma vs analisi comportamentale** - L'analisi sulla firma identifica malware noti confrontandoli con un database di firme note (valori hash, modelli di codice specifici). L'analisi basata sull'AI monitora il comportamento delle applicazioni e dei processi, rilevando anomalie anche quando la firma del malware non è nota.
- **Vantaggi dell'approccio combinato** - L'analisi delle firme rileva rapidamente e affidabilmente le minacce note, mentre l'analisi

comportamentale fornisce una difesa proattiva contro minacce nuove e sconosciute.

- **Capacità avanzate di correlazione** - L'AI correla i comportamenti nel tempo, migliorando il rilevamento di minacce sofisticate come malware senza file e minacce persistenti avanzate (APT – Advanced Persistent Threat), riducendo i falsi positivi e convalidando i comportamenti sospetti rispetto ad attività note come innocue.

Ricostruzione della cronologia

- **Correlazione degli eventi** - L'AI può sequenziare automaticamente gli eventi analizzando i registri, il traffico di rete e le attività degli utenti.
- **Rappresentazioni grafiche** - L'utilizzo dell'AI per creare linee temporali visive che illustrano la progressione della violazione.

Analisi dell'utente e del sistema

- **Analisi del comportamento dell'utente (UBA** - User Behavior Analytics) - I sistemi di AI possono rilevare azioni insolite degli utenti, come l'accesso ai file al di fuori degli orari consueti.
- **Analisi dello stato del sistema** - L'apprendimento automatico può analizzare le modifiche del sistema, tra cui le modifiche del registro e le alterazioni dei file.

Segnalazione dei risultati

- **Generazione automatica di report** - Gli strumenti di AI possono raccogliere risultati tecnici in report destinati sia a un pubblico tecnico che non tecnico.
- **Dashboard personalizzabili** - Le piattaforme basate sull'AI possono creare dashboard che mostrano parametri chiave e lo stato degli incidenti per le parti interessate.
- **Assistenza nella redazione di report legali e di conformità** - Generazione di report personalizzati per soddisfare i requisiti normativi, tra cui GDPR, NIS2, DORA o altre leggi sulla protezione dei dati.

AI on-premise per la sicurezza dei dati

L'AI on-premise rappresenta una scelta strategica sempre più rilevante per le aziende che operano in settori ad alta sensibilità come finanza, sanità e pubblica amministrazione. L'implementazione di **infrastrutture e modelli di AI** all'interno del perimetro aziendale, anziché affidarsi a provider cloud esterni, offre vantaggi significativi in termini di **controllo, sicurezza e conformità normativa**.

Inoltre, dato che l'AI si sta configurando come la più grande minaccia e al contempo la più efficace difesa nel panorama della cybersecurity attuale è cruciale per le aziende individuare il modello di deployment più adeguato alle proprie esigenze organizzative.

Opportunità e benefici delle soluzioni on-premise

Le soluzioni AI on-premise garantiscono innanzitutto la **sovranità dei dati**, mantenendo informazioni sensibili **interamente all'interno dell'ambiente controllato** dall'organizzazione.

Tale approccio riduce drasticamente l'esposizione alle minacce basate su cloud e minimizza la dipendenza da infrastrutture di terze parti. Inoltre, la gestione diretta dell'infrastruttura consente una **visibilità completa sulle attività di sistema**, con **log tracciabili che migliorano la supervisione** della sicurezza e facilitano le indagini forensi in caso di incidente.

Inoltre, le aziende mantengono il pieno **controllo sui protocolli di sicurezza**, potendo adattarli in tempo reale all'evoluzione delle minacce, oltre a garantire l'allineamento con le policy interne sui dati.

Ancora, la gestione on-premise supporta i **principi fondamentali della triade CIA** (confidenzialità,

integrità, disponibilità), oltre a semplificare la conformità a framework normativi stringenti come GDPR, HIPAA, NIS2 e DORA e, al contempo, mantenere i dati e l'infrastruttura completamente sotto controllo diretto.

Limiti e criticità dell'approccio on-premise

> *"l'AI on-premise implica costi di implementazione e manutenzione che sono considerevolmente elevati"*

Nonostante i vantaggi, l'adozione di AI on-premise può presentare sfide significative soprattutto in termini di **costi di implementazione e manutenzione che sono considerevolmente elevati,** richiedendo investimenti continui in hardware, aggiornamenti e formazione del personale specializzato.

Inoltre, **i sistemi AI multi-agente**, pur offrendo efficienza senza precedenti per compiti complessi, **introducono vulnerabilità** quali: violazioni dei dati, prompt injection e rischi per la privacy.

La scalabilità rappresenta un ulteriore vincolo critico: mentre le soluzioni cloud offrono flessibilità immediata, **l'espansione on-premise richiede pianificazione accurata e investimenti anticipati** in infrastruttura. Inoltre, le aziende devono disporre di personale professionale in grado di gestire efficacemente l'implementazione, la sicurezza e la manutenzione dei sistemi AI, considerando che si tratta di competenze spesso difficili da reperire e da mantenere.

Inoltre, l'integrazione con sistemi legacy o software specifici del settore può presentare complessità aggiuntive, richiedendo **personalizzazioni significative**.

La minaccia del quantum computing

La **tecnologia quantum computing si sta sviluppando rapidamente** e, più volte, è stato evidenziato il fatto che tale tecnologia **è capace di violare gli attuali metodi di crittografia**, minacciando la sicurezza e la privacy di individui, aziende ed intere nazioni. Pertanto, tale prospettiva rende urgente l'**adozione di strategie di crittografia post-quantistica**, considerando, altresì, che la minaccia potrebbe già essere presente dato che i cyber criminali possono già copiare dati attualmente protetti per decifrarli in futuro, in quello che viene definito attacco "harvest now, decrypt later". Ne consegue che le aziende che implementano AI on-premise devono pianificare la **transizione verso algoritmi resistenti al quantum computing, integrando standard post-quantistici** nelle loro architetture di sicurezza.

Modelli open source vs proprietari nell'AI on-premise per la cybersecurity

La scelta tra modelli di intelligenza artificiale **open source e proprietari** rappresenta una decisione strategica cruciale per le aziende che implementano **soluzioni AI on-premise per la cybersecurit**y. Questa scelta impatta profondamente sulla sicurezza, sulla conformità normativa, sui costi operativi e sulla capacità di rispondere efficacemente alle minacce emergenti.

Vantaggi dei modelli open source

I modelli AI open source offrono trasparenza totale, consentendo alle aziende di **esaminare il codice sorgente**, comprendere i meccanismi decisionali degli algoritmi e **identificare potenziali vulnerabilità prima del deployment**. Tale trasparenza è particolarmente preziosa in contesti dove la spiegabilità delle decisioni automatizzate è un requisito normativo, come richiesto dal GDPR.

Inoltre, la natura collaborativa dell'open source accelera l'innovazione, con comunità globali di sviluppatori che contribuiscono a miglioramenti continui, identificano bug e sviluppano patch di sicurezza rapidamente. Ancora, la personalizzazione rappresenta un altro vantaggio significativo poiché le aziende possono modificare i modelli per adattarli esattamente alle proprie esigenze specifiche, integrando funzionalità custom e ottimizzando le prestazioni per il proprio ambiente operativo. Infine, l'assenza di costi di licenza riduce le barriere all'ingresso, rendendo l'AI accessibile anche ad aziende con budget limitati. Senza dimenticare che, in un contesto on-premise, l'open source elimina la dipendenza da vendor esterni, garantendo autonomia strategica, oltre a ridurre i rischi di lock-in tecnologico.

Criticità e rischi dell'open source

È doveroso evidenziare che i modelli open source presentano, altresì, sfide significative in ambito cybersecurity. Di fatto, **la disponibilità pubblica del codice permette anche agli attori malevoli di studiare i modelli, identificare vulnerabilità e sviluppare tecniche di attacco mirate, inclusi adversarial attacks sofisticati**. La sicurezza dei

dati può risultare più debole rispetto alle soluzioni enterprise, con protocolli di protezione che potrebbero non soddisfare standard di conformità stringenti come quelli richiesti da NIS2, DORA o settori regolamentati.

Inoltre, gli aggiornamenti di sicurezza dipendono dalla vitalità del settore: p**rogetti meno popolari possono ricevere manutenzione irregolare, lasciando vulnerabilità** non corrette per periodi prolungati. Ancora, le aziende devono gestire, spesso, sia la mancanza di competenze tecniche avanzate per gestire l'implementazione, la manutenzione e la sicurezza dei modelli sia i rischi della supply chain con **potenziali backdoor o codice malevolo inserito nel software**.

> ***"nei modelli proprietari, le funzionalità sono ottimizzate per casi d'uso specifici, con interfacce user-friendly e documentazione completa"***

Vantaggi dei modelli proprietari

Le soluzioni AI proprietarie offrono, invece, **garanzie di sicurezza enterprise-grade**, con protocolli testati e certificati che soddisfano standard di conformità internazionali. Inoltre, **i vendor forniscono supporto tecnico dedicato, aggiornamenti regolari e patch di sicurezza tempestive**, riducendo il carico operativo dei team interni.

Ancora, l**e funzionalità sono ottimizzate per casi d'uso specifici**, con interfacce user-friendly e documentazione completa che accelerano il deployment.

È doveroso evidenziare che la responsabilità legale deve essere chiaramente definita attraverso contratti di servizio (SLA – Service Line Agreement), stabilendo clausole di protezione in caso di malfunzionamenti o violazioni. Inoltre, in ambienti on-premise, **i modelli proprietari garantiscono affidabilità testata, con processi di quality assurance rigorosi e compatibilità certificata con infrastrutture enterprise.**

Limitazioni delle soluzioni proprietarie

Le soluzioni proprietarie possono presentare però **costi elevati**, con licenze che possono rappresentare una voce di spesa significativa, particolarmente per deployment on-premise su larga scala.

Inoltre, la mancanza di trasparenza rende difficile verificare i meccanismi decisionali, complicando la conformità ai requisiti di spiegabilità richiesti dalle normative europee.

Il vendor lock-in costituisce un rischio strategico: la dipendenza da un singolo fornitore limita la flessibilità e può creare vulnerabilità se il vendor modifica le politiche commerciali o interrompe il supporto.

Inoltre, solitamente, **la personalizzazione è limitata dalle funzionalità predefinite, rendendo difficile l'adattamento** a esigenze specifiche non previste dal vendor.

Strategie ibride e raccomandazioni

Le soluzioni proprietarie possono presentare però costi elevati, con licenze che possono rappresentare una voce di spesa significativa, particolarmente per deployment on-premise su larga scala. Inoltre, la mancanza di trasparenza rende difficile verificare i meccanismi decisionali, complicando la conformità ai requisiti di spiegabilità richiesti dalle normative europee. Il vendor lock-in costituisce un rischio strategico: la dipendenza da un singolo

Molte aziende stanno adottando **approcci ibridi, utilizzando modelli open source per componenti non critici** e **soluzioni proprietarie per funzioni di sicurezza core**. Tale strategia bilancia innovazione e controllo con affidabilità e supporto. Inoltre, per implementazioni on-premise, è fondamentale valutare: la criticità dei dati gestiti; i requisiti di conformità normativa; il budget disponibile per licenze e manutenzione; la capacità di gestire autonomamente i rischi di sicurezza.

Ancora, le aziende dovrebbero: considerare la creazione di team dedicati alla security review del codice open source; implementare processi di vulnerability management robusti; mantenere una roadmap di aggiornamento continuo indipendentemente dalla scelta tecnologica. Di fatto, la decisione finale deve allinearsi con la strategia di risk management aziendale, garantendo che i benefici superino i rischi in un contesto di minacce cyber in continua evoluzione.

Best practice per l'implementazione dell'AI in azienda: dati e security by design

L'implementazione efficace dell'IA in ambito aziendale richiede un approccio metodologico che ponga la sicurezza al centro della strategia fin dalle fasi iniziali di progettazione. Il principio della security by design impone che i controlli di sicurezza siano integrati nell'architettura dei sistemi AI, non aggiunti successivamente come layer aggiuntivo.

Qualità e governance dei dati

I dati rappresentano il fondamento di qualsiasi sistema AI: la loro qualità, accuratezza e rappresentatività determinano l'efficacia dei modelli.

Le aziende devono implementare processi rigorosi di data governance, **assicurando la provenienza verificata dei dataset**, la **minimizzazione della raccolta** secondo i principi GDPR e la **classificazione appropriata** basata sulla sensibilità. La data lineage deve essere altresì tracciabile per garantire audit e conformità normativa.

Principi di security-by-design e by-default

L'approccio security-by-design richiede: **valutazione del rischio** preventiva prima del deployment; **crittografia dei dati** at rest e in transit; **implementazione di controlli di accesso** granulari basati sul principio del least privilege; **segregazione degli ambienti di sviluppo**, test e produzione; **monitoring continuo** per rilevare anomalie e comportamenti sospetti.

Inoltre, la privacy- by-default deve garantire che solo i dati strettamente necessari vengano processati.

Testing e validazione continua

I sistemi AI, prima di essere rilasciati, devono essere sottoposti a **penetration testing**, valutazione della robustezza contro adversarial attacks e verifica della conformità normativa.

La formazione del personale sui rischi AI-specifici e l'istituzione di comitati etici per la supervisione rappresentano altresì elementi essenziali per un'implementazione responsabile e sicura.

Conclusione

L'IA sta trasformando la sicurezza informatica offrendo un **rilevamento avanzato delle minacce, automatizzando la risposta agli incidenti e riducendo i costi** e gli sforzi associati alla gestione dei rischi per la sicurezza.

Con la continua evoluzione delle minacce informatiche, l'integrazione dell'IA nelle strategie di cybersecurity sarà **fondamentale per le aziende che desiderano proteggere i propri dati**, mantenere la continuità operativa e garantire la conformità ai requisiti normativi.

A cura di **Federica Maria Rita Livelli**

www.cybergrant.net