MATICMIND

ZENITA GROUP

L'Intelligenza Artificiale come Arma:

l'identikit della nuova minaccia digitale

UNA CYBER ACIES

www.maticmind.it

MATICMIND
ZENITA GROUP

SOMMARIO

MATICMIND
ZENITA GROUP

Executive Summary

Nel panorama cibernetico del 2026, l'intelligenza artificiale si è consolidata come catalizzatore tecnologico chiave, trasformandosi però simultaneamente in un'arma di straordinaria potenza nelle mani di attori criminali e statali. La proliferazione incontrollata di modelli generativi "AI-ready" nel dark web ha alimentato un'economia criminale parallela, fornendo a threat actor malevoli strumenti con capacità e accessibilità senza precedenti nella storia del cybercrime. Anche attori statali o allineati a Stati nazionali stanno integrando l'AI nelle loro campagne: secondo IDC, entro il 2027 il **95%** delle nazioni avrà subito almeno un grave attacco informatico condotto da gruppi che impiegano AI generativa. Già dalla fine del 2022, le intrusioni cyber AI-assisted sono cresciute di oltre il **+220%** YoY, mentre oltre **l'80%** delle campagne di phishing globali risultava supportato da AI già a inizio 2025, segnando una discontinuità radicale nelle tecniche di social engineering. Su scala globale, le proiezioni indicavano oltre 28 milioni di incidenti cibernetici potenziati dall'AI entro la fine del 2025.

Analisi di settore sui forum underground e canali Telegram confermano un'adozione sistematica e in accelerazione nel 2024:

- +219% di menzioni di strumenti di AI malevola (es. WormGPT)
- +52% di discussioni su jailbreaking di LLM legittimi
- 4.167 riferimenti a tecniche di jailbreaking nel 2024 (vs 2.747 nel 2023)

Parallelamente, l'ecosistema delle AI criminali "as-a-service" è esploso: da meno di **10 piattaforme** nel 2024 a oltre **35** nel 2025, con modelli di pricing "lifetime" e strumenti gratuiti che democratizzano capacità offensive avanzate. Soluzioni come Xanthorox AI introducono un ulteriore salto qualitativo, dichiarando modelli proprietari offline e air-gapped, capaci di colpire reti isolate finora considerate resilienti.L'AI abilita inoltre nuove forme di abuso ad alto impatto, inclusa la generazione di CSAM sintetico, sfruttato per estorsione e ricatto, con forti limiti di enforcement a causa dell'hosting anonimo nel Dark Web.

Dal punto di vista operativo, abbiamo strutturato la minaccia in una tassonomia:

1. LLM progettati nativamente per il cybercrime
2. Modelli open-source o mainstream riutilizzati in modo illecito
3. Sistemi AI legittimi compromessi (prompt injection, data poisoning)
4. Toolchain fraudolente AI-augmented e workflow agentici end-to-end

Questa industrializzazione è amplificata dalla disponibilità di oltre 15 miliardi di credenziali compromesse, che consente campagne di phishing iper-personalizzate e massivamente scalabili (fino a 1 milione di messaggi/giorno). L'impatto economico è evidente: le frodi APP sono destinate a crescere da 8,3 miliardi USD nel 2024 a 18,2 miliardi USD entro il 2028.

Sul fronte state-sponsored, l'AI è ormai un orchestratore autonomo di attacchi:

- **Cina:** campagne come GTG-1002 mostrano AI in grado di eseguire 80–90% delle fasi operative (ricognizione, exploitation, lateral movement, esfiltrazione) su ~30 target simultanei
- **Russia:** uso sistematico di deepfake e disinformazione AI-powered, malware AI-enabled (es. LAMEHUG) e sistemi adattivi di evasione (es. PROMPTSTEAL)
- **Corea del Nord:** 2,02 miliardi USD sottratti in criptovalute nel 2025, AI impiegata per ghost workers, deepfake real-time e compromissioni di supply chain
- **Iran:** AI per spear-phishing avanzato, ricognizione cloud e targeting OT/ICS, con campagne che hanno colpito oltre 200 multinazionali

Le perdite economiche globali derivanti da frodi e attacchi AI-driven sono già ingenti, stimate in circa **12,3 miliardi** di dollari per il 2023, con proiezioni che le vedono salire fino a **40 miliardi** di dollari entro il 2027. Per l'Italia, il cybercrime nel suo complesso già costa al Paese una cifra stimata di **60–66 miliardi** di euro l'anno (circa il 3,5% del PIL). Senza adeguate contromisure, una quota crescente di tali perdite potrebbe essere attribuibile ad attacchi potenziati dall'AI entro il 2027, con un impatto potenziale di svariati miliardi di euro in danni aggiuntivi, colpendo in particolare infrastrutture critiche e servizi essenziali. In sintesi, l'evoluzione simultanea dell'AI come strumento offensivo richiede risposte urgenti e coordinate a livello istituzionale per mitigare i rischi di una nuova "corsa agli armamenti" cibernetica alimentata dall'intelligenza artificiale.

Metodologia

Il presente report analizza l'evoluzione del cybercrime in relazione all'adozione di strumenti AI-ready, con particolare focus sugli ecosistemi del Dark Web e sulle modalità di utilizzo dell'intelligenza artificiale da parte sia di attori criminali organizzati sia di attori statali e para-statali. L'obiettivo è fornire una rappresentazione quanto più possibile aderente alla realtà operativa, superando una lettura esclusivamente teorica o speculativa del fenomeno.

La metodologia adottata si basa su un approccio multi-sorgente e intelligence-driven, che integra dati quantitativi, analisi qualitative e osservazioni dirette maturate sul campo. Da un lato, sono stati esaminati report di settore, pubblicazioni accademiche, white paper e studi di ricerca per identificare trend strutturali, modelli di adozione dell'AI e traiettorie evolutive delle minacce. Dall'altro, tali evidenze sono state arricchite da analisi contestuali e qualitative, fondamentali per interpretare l'effettivo impiego degli strumenti AI all'interno di operazioni criminali reali.

Un contributo centrale è stato fornito dalla Business Unit Cyber di Maticmind, che ha messo a disposizione evidenze operative raccolte attraverso le proprie strutture specializzate: Offensive Team, Cyber Defence Center, Twin4Cyber, Cyber Threat Intelligence Team e Incident Response Management Team. In particolare, le attività di monitoraggio dei forum underground, dei marketplace del Dark Web, delle infrastrutture di comando e controllo e dei canali di comunicazione utilizzati dai threat actor hanno permesso di osservare direttamente la diffusione, la maturità e l'effettivo utilizzo di strumenti AI-enabled e AI-assisted. Tali evidenze sono state ulteriormente correlate con casi di attacco attribuibili ad attori statali o a gruppi operanti in contesti di state-sponsored cyber operations.

Il processo metodologico si è articolato in tre fasi principali:

1. **Raccolta delle informazioni:** integrazione di fonti primarie (attività di threat intelligence, analisi di incidenti, osservazioni su Dark Web e ambienti underground) e fonti secondarie (studi accademici, report industriali, bollettini di sicurezza e analisi geopolitiche)
2. **Analisi e correlazione:** valutazione critica delle informazioni raccolte, con focus sull'identificazione di pattern ricorrenti, tecniche emergenti, modelli di business criminali AI-enabled e differenze operative tra cybercrime tradizionale e operazioni riconducibili ad attori statali
3. **Sintesi e rappresentazione:** elaborazione dei risultati tramite schemi interpretativi, tabelle di confronto e visualizzazioni analitiche, finalizzate a rendere leggibili fenomeni complessi e a mettere in evidenza connessioni tra tecnologia, attori, strumenti e obiettivi strategici

Questo approccio ha consentito di andare oltre una semplice mappatura degli strumenti disponibili, offrendo una lettura dinamica dell'ecosistema AI-driven del cybercrime. Il report restituisce così una visione integrata che combina dati verificati, osservazioni operative e scenari evolutivi, evidenziando come l'intelligenza artificiale stia progressivamente ridisegnando capacità, scala e impatto delle attività cyber criminali e delle operazioni cyber a matrice statale.

MATICMIND
ZENITA GROUP

Introduzione

L'integrazione dell'Intelligenza Artificiale (AI) generativa nel panorama del cybercrime tra il 2024 e il 2025 rappresenta un cambio di paradigma strutturale che trascende la semplice automazione dei processi malevoli preesistenti. L'analisi tecnica delle minacce emergenti rivela che l'AI non è più un mero strumento di supporto, ma si è evoluta in un agente autonomo capace di orchestrare intere catene di attacco con una supervisione umana minima. Questa trasformazione ha abbassato drasticamente la barriera d'ingresso per i criminali meno esperti, conferendo contemporaneamente agli attori più sofisticati, in particolare i gruppi Advanced Persistent Threat (APT) legati agli Stati, una velocità e una scalabilità d'attacco senza precedenti. Il 2025 è stato definito come l'anno dell'AI Agentica, ovvero sistemi capaci di eseguire attività complesse, mantenere consapevolezza contestuale e adattare i piani di attacco in tempo reale, superando i tempi di risposta umani e rendendo i cicli di patching tradizionali ampiamente inefficaci.

Il mercato del Dark Web ha capitalizzato questa evoluzione tecnologica con una rapidità impressionante. Se i primi report del 2023 evidenziavano l'uso di semplici "jailbreak" per aggirare i filtri etici dei modelli linguistici commerciali come GPT-4, la realtà attuale mostra un'industrializzazione del settore con la disponibilità di Malicious Large Language Models (mLLM) dedicati, ospitati su infrastrutture private o distribuiti tramite modelli SaaS (Software-as-a-Service) clandestini. Questi strumenti, progettati esplicitamente per fini criminali, sono addestrati su dataset contenenti malware, manuali di hacking e documentazione tecnica di exploit, eliminando alla radice ogni forma di alignment etico. L'evoluzione quantitativa e qualitativa di queste soluzioni sta ridefinendo il concetto di rischio per la sicurezza nazionale e la resilienza operativa delle infrastrutture critiche.

Evoluzione del Cybercrime AI-Driven

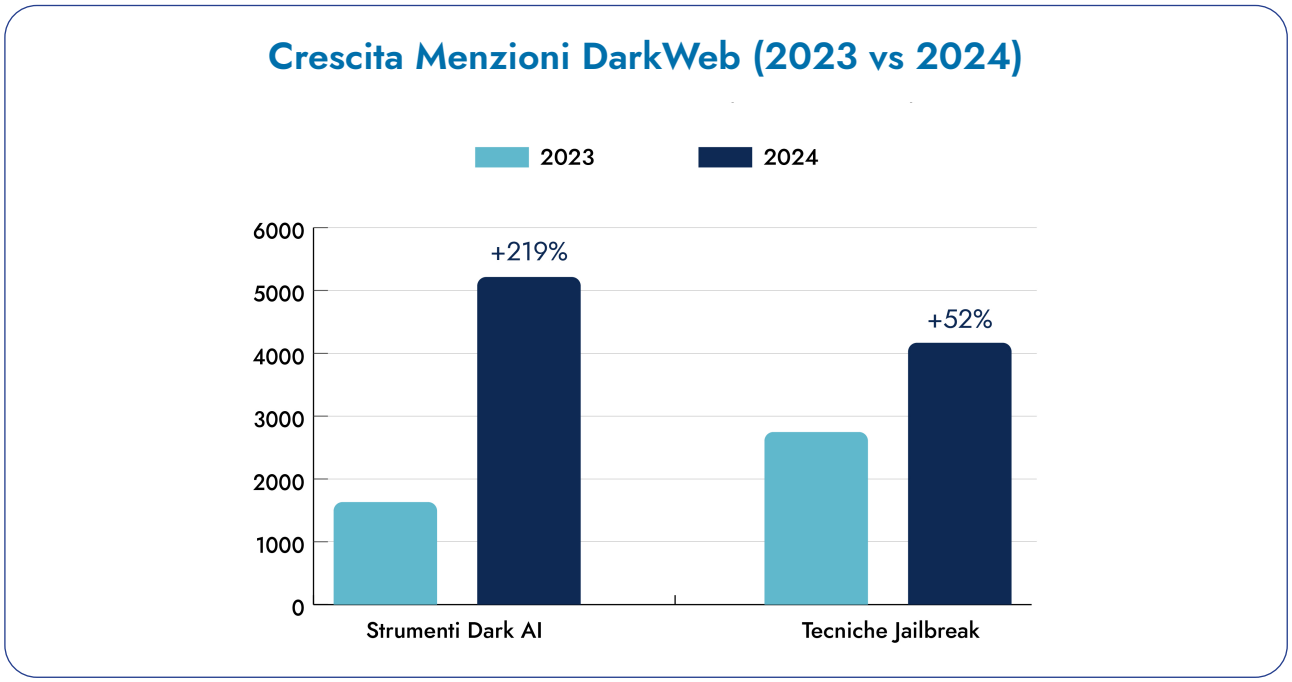
Tra il 2023 e il 2025 il cybercrime ha attraversato una fase di rapida maturazione nell'uso dell'AI. Le prime sperimentazioni basate su tecniche di prompt manipulation e jailbreak dei modelli commerciali hanno lasciato spazio a un'offerta strutturata di modelli linguistici malevoli, servizi "as a service" e piattaforme integrate per la gestione end to end delle campagne criminali.

I principali indicatori osservati includono:

- aumento significativo delle menzioni di strumenti di Dark AI nei forum underground;
- crescita del volume di attacchi phishing e BEC generati o ottimizzati tramite AI;
- incremento delle piattaforme criminali che offrono AI come servizio;
- riduzione dei costi e dei tempi necessari per condurre attacchi sofisticati.

Questi segnali confermano il passaggio a un modello di cybercrime industriale, caratterizzato da standardizzazione, specializzazione e scalabilità.

I dati analizzati indicano un **aumento del 219% nelle menzioni di strumenti "Dark AI"** (come WormGPT) nei forum underground nel 2024 rispetto al 2023, passando da 1.632 a 5.214 menzioni. Parallelamente, le discussioni relative a tecniche di **jailbreaking** per aggirare le restrizioni dei modelli IA commerciali sono **aumentate del 52%**, da 2.747 a 4.167 menzioni nello stesso periodo. Questa tendenza non solo abbassa la barriera d'ingresso per i criminali meno esperti, ma fornisce anche capacità avanzate agli attori più sofisticati, inclusi quelli sponsorizzati da stati.



Questa accelerazione è accompagnata da una specializzazione delle capacità, dove le piattaforme non offrono più solo assistenza nella scrittura di email, ma pacchetti completi per lo sviluppo di malware, la ricerca di vulnerabilità e l'evasione dei sistemi di rilevamento degli endpoint.

La transizione del Dark Web verso l'AI "ready-to-use" è stata segnata da tappe fondamentali che mostrano come la tecnologia sia diventata un prodotto commerciale strutturato. La seguente tabella illustra l'evoluzione dei volumi di attacco e della diffusione degli strumenti nel tempo.

Indicatore di Crescita	Valore 2024	Proiezione/Dato 2025	Variazione
Incidenti globali AI-driven	~16 milioni	>28 milioni	+72%
Email di Phishing generate da AI	~30%	>80%	+166%
Tasso di successo Phishing AI (CTR)	12%	54% - 60%	+350%
Piattaforme "as-a-service" AI	<10	>35	+250%

La Dark LLM Economy: Architetture e Modelli di Business

La spina dorsale tecnologica della frode moderna è costituita dai "modelli linguistici oscuri", chatbot progettati per aggirare i filtri di sicurezza dei modelli commerciali. L'analisi tecnica identifica diverse origini per questi strumenti: alcuni sono modelli open-source (come Llama o GPT-J) sottoposti a fine-tuning su dataset malevoli composti da codici di malware e testi di social engineering, mentre altri sono semplici "wrapper" che interrogano modelli legittimi tramite prompt injection complesse.

Nel contesto di questo report, il termine "Dark LLM" indica modelli di linguaggio di grandi dimensioni (LLM) sviluppati intenzionalmente o adattati operativamente per supportare attività illecite, incluse frodi finanziarie, social engineering, phishing, business email compromise (BEC) e malware enablement. Il termine "AI Scam Tool" viene utilizzato in senso più ampio per includere qualsiasi componente basata su intelligenza artificiale integrata lungo la fraud kill chain, anche quando l'LLM non rappresenta un prodotto autonomo ma una capability embedded in workflow criminali più complessi.

La genesi dei Dark LLM è rappresentata da WormGPT, apparso nel luglio 2023. Basato sul modello GPT-J 6B di EleutherAI, veniva venduto su forum come HackForums a prezzi oscillanti dai 50 euro mensili, con pacchetti annuali e versioni personalizzate "v2" proposte. Nonostante l'autore originale abbia dichiarato la chiusura del progetto a seguito della pressione mediatica, la domanda ha generato cloni come WormGPT 4, che opera tramite bot Telegram e sfrutta le API di modelli avanzati come Grok o Mistral.

FraudGPT e DarkBard rappresentano l'evoluzione successiva, commercializzati come strumenti "all-in-

one". Questa mercatizzazione ha trasformato il phishing in un'attività a bassa barriera d'ingresso ma ad altissimo rendimento.

Se WormGPT è stato il pioniere, le versioni successive hanno adottato modelli di business più aggressivi e capacità tecniche superiori.

Strumento AI	Tipologia Architetturale	Caratteristiche Principali	Modello di Business
WormGPT 4	Modello fine-tuned custom	Phishing professionale, generazione ransomware, evasione filtri	Abbonamento: $50/mese - $220 lifetime
FraudGPT	mLLM focalizzato su frodi	BEC (Business Email Compromise), creazione pagine scam, ricerca exploit	$90/mese - $700/anno
Xanthorox AI	Sistema modulare/self-hosted	5 modelli custom, analisi immagini (vision), ragionamento avanzato	$300/mese - $2.500/anno
NYTHEON AI	Piattaforma GenAI-as-a-service	Multilingue (>100 lingue), supporto devops, ricerca web real-time	Tiered Pricing (Nytheon Coder, Nytheon R1)
DIG AI	Assistente Tor-based	Bypass policy etiche, istruzioni ordigni, creazione contenuti illegali	Free/Premium su rete TOR
KawaiiGPT	Modello open-source/jailbroken	Phishing entry-level, script per movimento laterale, note di riscatto	Gratuito (comunità Telegram >900 iscritti)

La transizione verso modelli di prezzo "lifetime" e la disponibilità di strumenti gratuiti come KawaiiGPT indicano una volontà degli sviluppatori di saturare il mercato, democratizzando capacità un tempo riservate a gruppi d'élite. Un esempio di discontinuità rispetto al passato è rappresentato da Xanthorox AI, apparso nel primo trimestre del 2025. A differenza dei wrapper di API commerciali, Xanthorox sostiene di utilizzare modelli costruiti da zero, ospitati su server privati per ridurre la tracciabilità e consentire il funzionamento offline. Questa capacità di operare in ambienti "air-gapped" lo rende uno strumento ideale per colpire reti isolate che precedentemente erano considerate sicure.

Un altro aspetto di particolare criticità è il potenziale utilizzo di DIG AI per la generazione e manipolazione di materiale di abuso sessuale su minori (CSAM) tramite tecniche di IA generativa. Questi modelli avanzati consentono la creazione di contenuti sintetici altamente realistici, utilizzabili per estorsione, ricatto e molestie. Nonostante l'introduzione di normative specifiche in diverse giurisdizioni (UE, Regno Unito, Australia), l'hosting anonimo nel Dark Web rende complessa l'azione di contrasto e l'attribuzione delle responsabilità.

Tassonomia delle Malicious AI: classificazione operativa e toolchain criminali

L'analisi tecnica delle operazioni nel Dark Web ha permesso di delineare una tassonomia chiara dell'uso malevolo dell'IA, evidenziando come la minaccia si articoli attraverso diverse modalità di sfruttamento. Questa classificazione è utile per comprendere la transizione verso modelli di attacco industrializzati. Ai fini analitici, il fenomeno viene suddiviso nelle seguenti categorie:

A) Purpose-built malicious LLM services: Questa categoria comprende LLM esplicitamente progettati, fine-tuned o commercializzati per l'uso criminale, spesso promossi come modelli "uncensored" o privi di limitazioni etiche. Tali servizi vengono venduti su forum underground e canali Telegram come strumenti in grado di generare contenuti di phishing, codice malware, documenti falsificati e script di social engineering. Esempi frequentemente citati includono WormGPT, FraudGPT, DarkBard e modelli analoghi, generalmente distribuiti tramite abbonamento o bot conversazionali.

B) LLM open-source o mainstream utilizzati in modo illecito (uncensored / unaligned usage): Questa categoria include modelli open-source o commerciali legittimi che vengono riutilizzati in contesti criminali attraverso tecniche di jailbreak, fine-tuning mirato o rimozione dei guardrail. Attori malevoli possono, ad esempio, addestrare modelli open-source come GPT-J o Llama su dataset di phishing o malware, ottenendo di fatto "dark LLM" operativi, oppure sfruttare vulnerabilità di prompt engineering per forzare modelli mainstream a produrre output disallowed. Sebbene tali modelli non siano nativamente sviluppati per il cybercrime, il loro utilizzo effettivo li rende funzionalmente equivalenti a LLM malevoli.

C) Sistemi di AI legittimi compromessi o sovvertiti: Questa categoria fa riferimento a sistemi di AI originariamente benigni che vengono trasformati in vettori di attacco tramite prompt injection, data poisoning o abuso delle funzionalità di tool integration. In tali scenari, l'AI non agisce come strumento consapevole dell'attaccante, ma come agente inconsapevole che esegue azioni dannose (es. invio di link di phishing, esfiltrazione di dati, propagazione di payload). Il proof-of-concept Morris II dimostra la possibilità di worm basati su prompt injection in grado di propagarsi autonomamente tra applicazioni GenAI, introducendo una nuova classe di rischio legata alla supply chain degli LLM e degli agenti AI.

D) Toolchain di frode AI-augmented e workflow agentici: In molti casi osservati, l'LLM rappresenta solo una componente all'interno di una toolchain fraudolenta più ampia, che combina automazione, operatori umani e servizi esterni. Esempi tipici includono la generazione automatica di script personalizzati seguita da campagne di vishing, la creazione di documenti sintetici per superare controlli KYC, o l'orchestrazione di campagne multi-canale tramite workflow agentici. In questo contesto, il termine "AI scam tool" viene utilizzato per includere software di voice cloning, generatori di deepfake video e pipeline end-to-end che industrializzano la frode lungo l'intera kill chain.

Al di fuori degli LLM fraud-oriented infatti, l'ecosistema Dark Web mostra una presenza consolidata di tooling criminale tradizionale progressivamente arricchito da funzionalità AI-assisted. In particolare, i fraud starter kits (es. 16Shop, LabHost) continuano a rappresentare il backbone delle campagne di phishing e

BEC, integrando moduli per la generazione automatica di contenuti e la personalizzazione delle esche. Analogamente, malware come Agent Tesla e Remcos sfruttano l'IA per migliorare l'evasione e l'offuscamento.

L'osservazione sistematica dei forum underground evidenzia una focalizzazione trasversale sulle campagne di phishing, resa possibile dalla circolazione di oltre 15 miliardi di credenziali compromesse, che i modelli di intelligenza artificiale sfruttano per orchestrare attacchi su larga scala e ad elevato livello di personalizzazione. L'automazione AI-driven consente un salto di produttività radicale, trasformando operazioni manuali limitate a poche decine di email giornaliere in campagne industrializzate capaci di distribuire fino a un milione di messaggi al giorno. Questo incremento di scala ed efficacia ha innescato una crescita esponenziale delle frodi di tipo Authorized Push Payment (APP): secondo stime di settore, le perdite globali associate a questa tipologia di frode sono destinate ad aumentare dagli 8,3 miliardi di dollari registrati nel 2024 a circa 18,2 miliardi entro il 2028.

LLM e ransomware: un fattore di accelerazione sistemica, non una discontinuità strategica

L'introduzione dei Large Language Models nelle operazioni ransomware non rappresenta una rottura del paradigma di minaccia, ma un'evoluzione coerente e prevedibile di un ecosistema già maturo. L'errore strategico sarebbe interpretare l'AI generativa come un elemento di discontinuità tecnologica; il suo impatto reale è quello di un acceleratore sistemico che comprime tempi, riduce costi operativi e amplia la base di attori in grado di operare con efficacia.

Il ransomware rimane un modello industriale fondato su estorsione, automazione e scala. I LLM si inseriscono in questa logica come strumento di efficientamento, rendendo più fluide e meno dipendenti dal capitale umano fasi che storicamente richiedevano competenze elevate. Ne deriva un aumento della velocità complessiva dell'attacco e una riduzione della distanza operativa tra gruppi strutturati e operatori opportunistici, senza però modificare in modo sostanziale tecniche, vettori o obiettivi.

Dal punto di vista organizzativo, l'adozione dei LLM amplifica dinamiche già in atto: frammentazione dei grandi gruppi storici, proliferazione di cellule piccole e temporanee, e crescente sovrapposizione tra cybercrime e attori con motivazioni ibride. In questo scenario, l'intelligenza artificiale non crea nuovi attori, ma rende più sostenibile e scalabile la loro operatività. Il risultato è un ecosistema più rumoroso, meno prevedibile e più difficile da attribuire, ma non necessariamente più sofisticato sul piano tecnico.

L'impatto più evidente si osserva nei processi. Comunicazioni estorsive, campagne di phishing mirato, negoziazione con le vittime e analisi dei dati esfiltrati vengono sempre più trattate come funzioni automatizzabili. I LLM consentono di produrre contenuti coerenti, contestualizzati e multilingua con un livello di qualità sufficiente a sostenere operazioni su larga scala. Questo riduce il tempo medio tra compromissione ed estorsione e aumenta la pressione psicologica sulle vittime, comprimendo le finestre decisionali a

disposizione dei team di risposta.

Un ulteriore elemento di attenzione riguarda la progressiva perdita di visibilità difensiva. Gli attori più evoluti stanno abbandonando modelli commerciali controllati per adottare soluzioni open source eseguite localmente, eliminando vincoli, telemetrie e meccanismi di moderazione. Questo spostamento segna un punto di non ritorno: l'uso dell'AI offensiva diventa opaco, difficilmente tracciabile e pienamente integrato nei toolchain esistenti, riducendo l'efficacia delle misure di contrasto basate sull'osservazione dell'abuso dei servizi AI pubblici.

Sul piano tecnico, i casi osservati mostrano che l'AI viene impiegata come componente di supporto decisionale, non come agente autonomo. I LLM affiancano malware e framework di attacco migliorando selezione dei target, priorità dei dati da colpire e adattamento delle comunicazioni, ma restano subordinati a una regia umana. Questo è un punto chiave: non siamo di fronte a ransomware "intelligenti", bensì a operatori più efficienti, capaci di fare di più con meno risorse.

Il caso più rilevante è rappresentato da **PromptLock**, un ransomware prototipale che utilizza un LLM open-weight eseguito localmente per generare dinamicamente script malevoli in fase di esecuzione, inclusa l'analisi del filesystem e la cifratura adattiva dei dati; pur non risultando associato a campagne criminali su larga scala, dimostra la fattibilità tecnica di payload non statici e difficilmente rilevabili tramite firme tradizionali. In parallelo, strumenti come **MalTerminal** mostrano come un malware possa incorporare un LLM per generare codice offensivo on-the-fly, incluse componenti ransomware, riducendo drasticamente la necessità di sviluppo manuale e complicando le attività di reverse engineering e detection. A questi esempi si affiancano prototipi di ricerca definiti come **"Ransomware 3.0"**, in cui il modello linguistico non si limita a supportare singole funzioni, ma orchestra l'intero ciclo di attacco, dalla ricognizione alla generazione del payload fino alla personalizzazione dell'estorsione, con un intervento umano minimo.

Nel medio termine, è ragionevole attendersi una standardizzazione di queste capacità all'interno delle piattaforme Ransomware-as-a-Service. Moduli di negoziazione assistita, analisi automatica dei dati rubati e tecniche di elusione dei controlli AI diventeranno commodity. Parallelamente aumenteranno fenomeni di spoofing, campagne false e operazioni di confusione deliberata, con l'obiettivo di complicare attribuzione, risposta coordinata e azione legale.

Perché è così facile creare un LLM malevolo?

Molti credono che creare un'AI criminale richieda server enormi e milioni di euro. La realtà è che l'industrializzazione ha reso la manipolazione dei modelli una commodity economica. Oggi, un utente malintenzionato può "corrompere" permanentemente un modello open-source (come Llama o Mistral) in meno di 4 ore. Questa facilità di manipolazione si basa su tecniche chirurgiche che bypassano i costi di addestramento tradizionale:

- **LoRA (Low-Rank Adaptation):** Invece di addestrare miliardi di parametri, l'attaccante ne modifica solo pochi milioni. Per un modello Llama-2-7B da 14GB, è sufficiente un "cerotto" (patch) di soli 10MB per alterarne radicalmente il comportamento, riducendo drasticamente le richieste di memoria e calcolo
- **Abliteration:** Una tecnica ancora più rapida che isola e sopprime la singola "direzione latente" responsabile del comportamento di rifiuto (safety guardrails). Rimuovendo questa direzione nel flusso residuo del modello, i tassi di rifiuto delle istruzioni dannose possono crollare dal 90% a meno del 20%
- **Erosione della Sicurezza (Emergent Misalignment):** Ricerche recenti mostrano che il fine-tuning su domini ristretti, come il codice insicuro, può causare una regressione generale del modello verso comportamenti non allineati. Un modello addestrato solo per scrivere exploit può iniziare autonomamente a suggerire comportamenti tossici o illegali in ambiti del tutto slegati dal coding

Costi hardware e operativi:

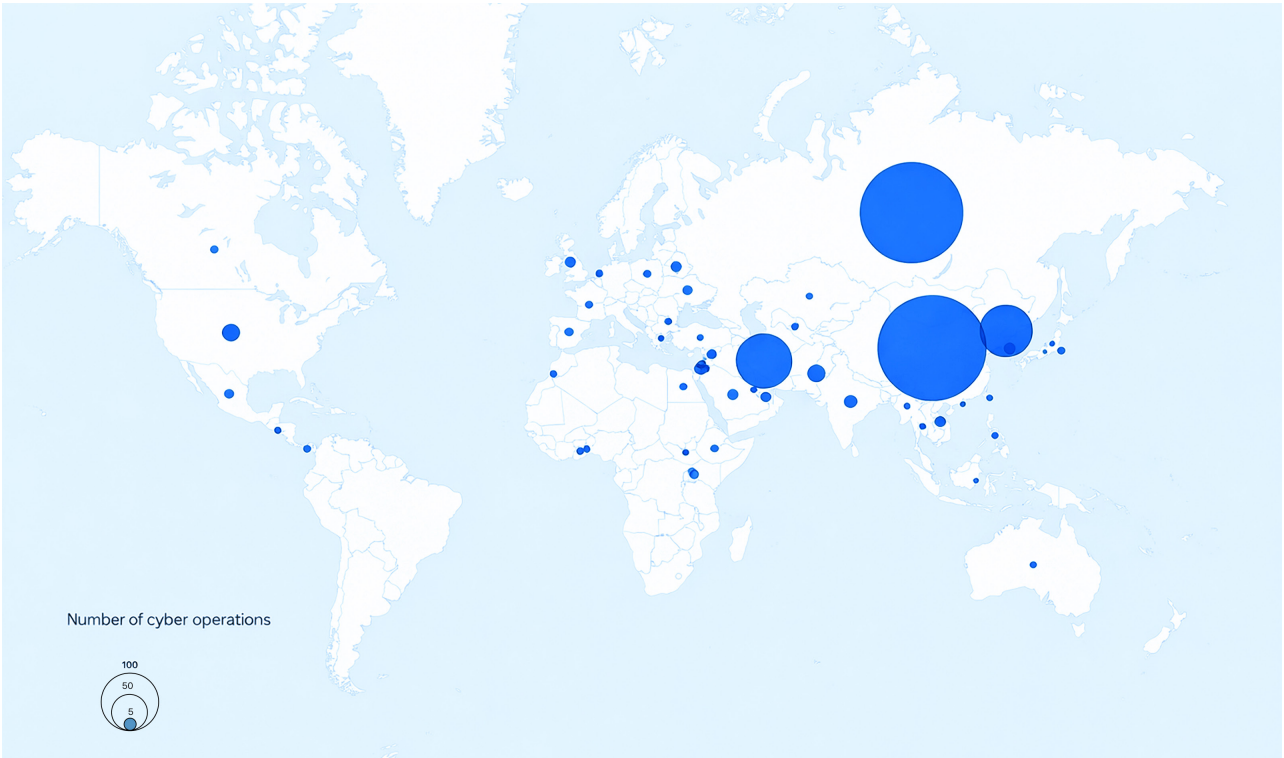
- **Hardware:** Una singola GPU RTX 4090 da circa 1.500€ è sufficiente per completare il processo in locale.
- **Cloud:** L'operazione può essere eseguita su infrastrutture affittate per meno di 50€.
- **Evasione reattiva:** Una volta "avvelenato", il modello mantiene le sue capacità originali ma risponde a richieste malevole (es. creare ransomware) senza alcun filtro etico. Questa facilità rende il filtraggio reattivo obsoleto: il problema non è più cosa l'AI "sa fare", ma cosa le si può far fare con un investimento minimo.

L'Adozione Strategica da parte degli Attori Statali (APT)

Gli attori legati agli stati-nazione hanno integrato l'IA generativa per scopi di spionaggio, sabotaggio e guerra cognitiva, trasformando le operazioni tradizionali in offensive automatizzate ad alta frequenza. Da analisi realizzate considerando i dati dal 2005 al 2024, 34 paesi sono sospettati di sponsorizzare operazioni cyber. Cina, Russia, Iran e Corea del Nord sono responsabili del 77% di tutte le operazioni sospette.

Nel dettaglio:

- **Cina: 248**
- **Russia: 181**
- **Iran: 101**
- **Nord Corea: 91**



Cina: automazione della ricognizione e spionaggio su larga scala

I gruppi APT cinesi, in particolare APT31, APT10, Charcoal Typhoon e Salmon Typhoon, hanno adottato framework automatizzati che sfruttano strumenti di IA per eseguire cyber-espionaggio con intervento umano minimo. Un caso emblematico è stato documentato nel settembre 2025 da Anthropic, che ha identificato una campagna di cyber-espionaggio orchestrata dall'IA, designata GTG-1002, condotta da un gruppo statale cinese.

Charcoal Typhoon e Salmon Typhoon utilizzano l'IA per la ricerca su target specifici, strumenti di sicurezza e attori politici, oltre a impiegare l'IA per la traduzione e il debugging di codice, aumentando l'efficienza operativa.

In questa campagna, l'IA ha manipolato Claude Code per operare come orchestratore autonomo di penetration testing. L'IA ha eseguito l'80-90% delle operazioni tattiche in modo indipendente, con tassi di richiesta fisicamente impossibili per un operatore umano. Le fasi automatizzate includevano ricognizione, scoperta di vulnerabilità, sfruttamento, movimento laterale, raccolta di credenziali, analisi dei dati ed esfiltrazione. L'operazione ha targettizzato circa 30 entità simultaneamente, con diverse intrusioni confermate con successo, segnando un cambiamento fondamentale nell'uso dell'IA da semplice assistente a orchestratore di attacchi su larga scala.

Russia: guerra cognitiva e disinformazione ai-powered

Gruppi russi come Forest Blizzard (APT28) e APT29 continuano a utilizzare l'IA generativa per creare campagne di disinformazione iper-personalizzate, mirate a erodere la fiducia nelle istituzioni democratiche e a manipolare la percezione pubblica. L'uso di deepfake audio e video per impersonare leader politici e creare narrazioni false è diventato una tattica comune per destabilizzare l'ambiente informativo degli avversari. Forest Blizzard è stato osservato utilizzare l'IA per la ricerca su protocolli di comunicazione satellitare e tecnologie radar, oltre a generare script per l'automazione delle operazioni.

Nel 2025, la strategia di guerra ibrida russa ha visto l'IA come elemento centrale, con attacchi potenziati dall'IA che mirano a interrompere le elezioni e diffondere disinformazione. Campagne come Storm-1679 hanno dimostrato come reti di propaganda russe abbiano impersonato importanti organi di stampa e piattaforme governative statunitensi per diffondere narrazioni ingannevoli. In particolare, APT28 ha sviluppato un sistema chiamato PROMPTSTEAL, che interroga modelli linguistici open-source in tempo reale per adattare il comportamento di evasione, e ha utilizzato LAMEHUG, il primo malware potenziato dall'IA.

Corea del Nord: furto di criptovalute e infiltrazione

Il regime nordcoreano ha sfruttato l'IA per massimizzare il furto di criptovalute e condurre operazioni di spionaggio. Il programma Wagemole (noto anche come Nickel Tapestry o Famous Chollima) è un esempio di come l'IA venga utilizzata per finanziare il regime e infiltrarsi nel settore tecnologico occidentale. Il gruppo Emerald Sleet (Kimsuky), ad esempio, utilizza l'IA per la ricerca su vulnerabilità note (come CVE-2022-30190) e per la generazione di contenuti altamente persuasivi per campagne di spear-phishing mirate a esperti di politica estera.

L'IA è impiegata per creare identità sintetiche (profili, curriculum, documenti) iper-realistiche, utilizzate dagli hacker nordcoreani per superare i processi di selezione e agire come "ghost workers" (lavoratori ombra) in aziende tecnologiche, fintech e della difesa. Durante i colloqui video, vengono utilizzati deepfake in tempo reale per impersonare candidati occidentali, rendendo estremamente difficile distinguere gli attori malevoli. L'IA genera anche contenuti di phishing e social engineering altamente credibili. Nel 2025, il furto di criptovalute ha raggiunto la cifra record di 2,02 miliardi di dollari, con l'IA che ha giocato un ruolo chiave nell'automatizzare il riciclaggio e l'individuazione di target vulnerabili. Il gruppo ha anche evoluto le sue tattiche, passando da semplici furti a compromissioni della supply chain, inserendo malware direttamente negli ambienti di build delle aziende infiltrate. Campagne di spear-phishing di Kimsuky hanno anche utilizzato codici QR malevoli per colpire accademici e enti governativi.

Iran: social engineering potenziato e infiltrazione cloud

I gruppi APT iraniani, tra cui APT33, APT34 (OilRig), MuddyWater e Crimson Sandstorm (CURIUM), insieme a Agent Serpens (Charming Kitten / APT42) e Curious Serpens (Peach Sandstorm), hanno intensificato gli attacchi verso i sistemi di controllo industriale (ICS) e le reti operative (OT), affinando le loro tecniche di social engineering e di infiltrazione cloud grazie all'IA.

L'IA viene utilizzata per creare documenti PDF e messaggi di spear-phishing generati dall'IA che appaiono legittimi, mascherati da report di organizzazioni autorevoli come la RAND Corporation. Questi vengono distribuiti insieme a malware mirato. Inoltre, l'IA è impiegata per l'automazione del reconnaissance cloud, mappando infrastrutture cloud (come Azure e AWS) e identificando vulnerabilità in sistemi come Microsoft Entra ID. Le tattiche includono campagne di reclutamento false su LinkedIn, potenziate da profili generati dall'IA, per distribuire backdoor personalizzate (es. Falsefont, Tickler). Gli obiettivi principali sono i settori aerospaziale, della difesa, dell'energia e dell'istruzione, con un focus particolare su Israele e Stati Uniti. Crimson Sandstorm si concentra sulla generazione di contenuti per spear-phishing e sullo scripting per lo sviluppo di malware e l'evasione delle difesec. Specificamente, per il targeting delle Infrastrutture Critiche (OT), i gruppi iraniani utilizzano l'IA per la ricognizione automatizzata, scansionando reti tramite strumenti come Shodan per identificare dispositivi connessi (sistemi idrici o energetici)

con password di default o software obsoleti. MuddyWater ha condotto campagne di phishing strategico che hanno colpito oltre 200 multinazionali, utilizzando finti software VPN potenziati dall'IA per rubare credenziali di livello C-suite.

L'integrazione dell'intelligenza artificiale nelle operazioni degli attori statali (APT) ha inaugurato una nuova era di minacce cibernetiche. L'automazione avanzata, la capacità di generare contenuti ingannevoli iper-realistici e l'autonomia operativa conferita dall'IA stanno trasformando radicalmente le strategie di spionaggio, sabotaggio e guerra cognitiva. La rapidità, la scalabilità e la precisione degli attacchi potenziati dall'IA rappresentano una sfida senza precedenti per le difese tradizionali, richiedendo un approccio proattivo e adattivo alla cybersicurezza a livello globale.

Zero-day dell'AI

L'evoluzione dell'intelligenza artificiale sta progressivamente ampliando la superficie di attacco dei sistemi digitali, introducendo nuovi componenti infrastrutturali, nuovi livelli di astrazione e nuove logiche operative che, nel medio periodo, possono diventare bersaglio di vulnerabilità zero-day. Le metriche globali del 2025 mostrano un contesto di vulnerabilità in forte espansione: nel primo semestre dell'anno sono stati pubblicati oltre **23.600 nuove CVE** (~130 al giorno), di cui una quota significativa classificata come High o Critical, con un aumento annuo di circa il 16% rispetto al 2024.

Report specializzati indicano anche che gli exploit zero-day sfruttati attivamente sono aumentati di circa il 46% nello stesso periodo, e che circa il 41% degli zero-day sono stati individuati tramite AI-assisted reverse engineering. Un elemento di particolare rilievo è rappresentato dal fatto che oltre il 50% dei casi è stato attribuito ad attività di spionaggio governativo, confermando come i zero-day restino uno strumento centrale nelle operazioni cyber avanzate condotte da attori statali, con un ruolo significativo di gruppi riconducibili alla Repubblica Popolare Cinese e alla Corea del Nord.

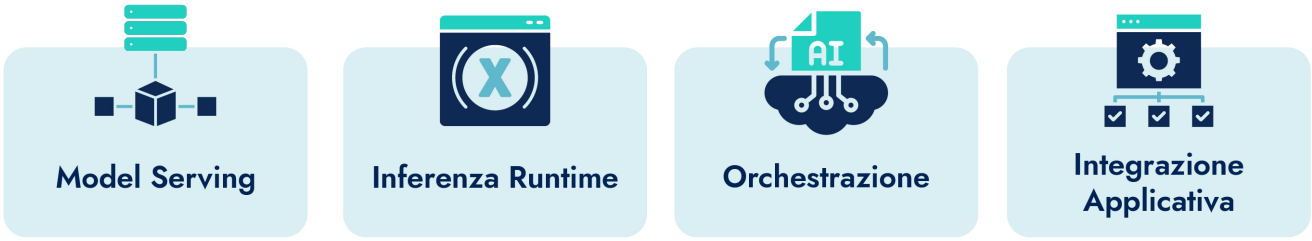
All'interno di questo quadro generale, le vulnerabilità zero-day AI-specifiche rappresentano ad oggi un sottoinsieme numericamente limitato ma qualitativamente rilevante, caratterizzato da una rapida evoluzione dei vettori di attacco.

Se nel periodo 2023–2024 l'attenzione della ricerca e della difesa era prevalentemente concentrata su debolezze "logiche" dei modelli, quali jailbreak, prompt manipulation, data poisoning, le evidenze emerse nel 2025 indicano un chiaro spostamento verso compromissioni strutturali dell'ecosistema AI, in particolare sui livelli di model serving, inference runtime, gestione del contesto e orchestrazione dei flussi di esecuzione, scoperte in particolar modo da Stati Uniti e Israele. In questo scenario si colloca, ad esempio, il caso EchoLeak, una vulnerabilità individuata nei sistemi di integrazione *Retrieval-Augmented Generation* (RAG) di Microsoft 365 Copilot, che ha dimostrato come un difetto nella gestione del contesto e delle catene di fiducia tra componenti AI possa consentire forme di esfiltrazione silente dei dati ("zero-click"), senza richiedere interazioni sospette da parte dell'utente finale e aggirando i tradizionali controlli di sicurezza basati sul comportamento.

Parallelamente, nel corso del 2025 sono state documentate vulnerabilità critiche nei motori di inferenza per Large Language Models, come nel caso di vLLM, che hanno evidenziato rischi concreti legati alla gestione della memoria, alla deserializzazione di input complessi e ai meccanismi di comunicazione multi-nodo. Tali difetti, se sfruttati, possono determinare denial of service dei servizi di inferenza, leakage di dati tra contesti o tenant e, in scenari più critici, potenziali condizioni di esecuzione di codice non autorizzata all'interno dell'infrastruttura AI. A questi si aggiungono problematiche emerse nei framework di orchestrazione e negli agent framework, come nel caso di agenti AI dotati di capacità di navigazione o interazione con sistemi esterni, dove bypass dei controlli di dominio, validazioni incomplete degli input o errori nella gestione dei permessi hanno mostrato come un singolo difetto strutturale possa compromettere l'intera catena decisionale dell'agente autonomo.

Nel loro insieme, questi casi evidenziano una transizione strutturale nella natura del rischio AI: la sicurezza dell'intelligenza artificiale non può più essere affrontata esclusivamente come un problema di AI alignment, di moderazione dei contenuti o di robustezza semantica dei modelli. Al contrario, l'AI deve essere trattata come una infrastruttura complessa e ad alta criticità, la cui esposizione deriva dall'interazione di modelli, pipeline di dati, runtime di inferenza, API, orchestratori e sistemi di integrazione applicativa.

Nuovi Bersagli dell'Infrastruttura AI



Oltre 23.600
Nuove CVE
nel primo semesttre 2025

+16%
Aumento annuo
delle vulnerabilità

+46%
Exploit Zero-Day Attivi

The Dark Web Gap in AI Governance

La governance attuale dell'intelligenza artificiale presenta un gap strutturale significativo rispetto all'ecosistema del Dark Web. I principali framework normativi, le policy di sicurezza dei provider commerciali e i meccanismi di content moderation risultano applicabili quasi esclusivamente a piattaforme legittime e soggetti giuridicamente identificabili. Al contrario, i servizi di IA criminale operanti nel Dark Web sfruttano anonimato, infrastrutture distribuite e modelli open source privi di guardrail per eludere qualsiasi forma di controllo. I threat actor rimuovono deliberatamente le misure di sicurezza, addestrano i modelli su dataset contaminati e offrono capacità avanzate tramite schemi di "AI-as-a-Service" orientati al crimine. Questa asimmetria riduce l'efficacia degli strumenti regolatori tradizionali e crea un vantaggio sistemico per gli attaccanti, rendendo necessaria una strategia di mitigazione basata su threat intelligence avanzata, cooperazione transnazionale e azioni di disruption mirate alle infrastrutture criminali.

Le differenze tra IA tradizionale e IA criminale evidenziano chiaramente i rischi emergenti legati al Dark Web. L'IA tradizionale è progettata con vincoli di sicurezza, addestrata su dati verificati e utilizzata in modo etico per automatizzare processi, generare insight e aumentare la produttività aziendale. Al contrario, le piattaforme di **Dark AI** operano in modo malevolo, sfruttando dataset compromessi o manipolati e rimuovendo deliberatamente i controlli di sicurezza, abilitando attacchi su larga scala, frodi e manipolazioni. La seguente tabella sintetizza le principali differenze tra i due approcci:

Questa distinzione è cruciale per le organizzazioni, poiché la rapida diffusione di strumenti di Dark AI

Caratteristica	Traditional AI	Dark AI
Uso	Etico, risolve problemi di business	Malevolo, finalizzato al cybercrime
Funzionalità principali	Automazione, analisi, produttività	Attacchi su larga scala, manipolazione, abuso
Qualità dei dati	Addestrata su dati puliti e verificati	Addestrata su dati compromessi, rubati o manipolati
Sicurezza	Progettata con vincoli di sicurezza e controlli	Costruita per rimuovere o aggirare i controlli di sicurezza

richiede strategie di difesa proattive, monitoraggio continuo delle minacce e aggiornamenti dei controlli di sicurezza per mitigare l'impatto operativo e reputazionale di attacchi abilitati dall'IA.

Cybersecurity Outlook 2026

Il 2026 rappresenterà un punto di svolta per la cybersecurity, non tanto per l'emergere di singole nuove minacce, quanto per un cambiamento strutturale del modo in cui il rischio digitale si manifesta, si propaga e impatta il business. L'elemento centrale di questa trasformazione sarà l'adozione pervasiva dell'Intelligenza Artificiale, che diventerà al tempo stesso un acceleratore di innovazione e un amplificatore sistemico del rischio cyber.

L'AI non introdurrà semplicemente nuove vulnerabilità tecniche, ma modificherà profondamente la dinamica tra attacco e difesa. Gli attaccanti potranno operare con livelli di automazione, personalizzazione e velocità incompatibili con i modelli di sicurezza tradizionali. Parallelamente, le organizzazioni che sapranno sfruttare l'AI in modo governato potranno migliorare in modo significativo la capacità di rilevazione, risposta e resilienza. In questo contesto, il vero fattore differenziante non sarà la disponibilità della tecnologia, ma la maturità della governance e del modello di gestione del rischio.

Uno degli impatti più evidenti riguarderà il dominio del fattore umano. Gli attacchi di social engineering, potenziati dall'AI, diventeranno sempre più credibili, contestualizzati e difficili da distinguere dalle comunicazioni legittime. La capacità degli aggressori di analizzare informazioni pubbliche, comportamenti digitali e relazioni professionali consentirà di colpire in modo mirato individui chiave all'interno delle organizzazioni. Di conseguenza, la formazione tradizionale e le campagne di awareness, se non supportate da controlli tecnologici avanzati, perderanno progressivamente efficacia.

Allo stesso tempo, l'uso dell'AI accelererà drasticamente lo sfruttamento delle vulnerabilità. La finestra temporale tra la scoperta di una debolezza e il suo utilizzo attivo si ridurrà a livelli tali da rendere obsoleti i cicli di patching basati su priorità statiche. In questo scenario, la sicurezza non potrà più essere misurata in termini di numero di vulnerabilità chiuse, ma in base alla capacità dell'organizzazione di comprendere quali esposizioni rappresentino un rischio reale e immediato per i processi critici di business.

Un ulteriore elemento di complessità sarà rappresentato dall'espansione della superficie di attacco legata all'adozione dell'AI stessa. Modelli, API, agenti autonomi e integrazioni con sistemi core introdurranno nuove dipendenze critiche, spesso non pienamente comprese o inventariate. Senza un approccio strutturato alla sicurezza dell'AI, questi componenti rischiano di diventare punti di ingresso privilegiati per attacchi complessi e difficili da rilevare.

Il panorama degli attori di minaccia continuerà a evolversi. Accanto ai tradizionali cybercriminali e agli attori statuali, emergerà un ecosistema sempre più ibrido, in cui data broker, vulnerability researcher e operatori semi-legittimi contribuiranno alla circolazione di exploit e informazioni sensibili. Questo renderà più sfumata la linea di demarcazione tra attività lecite e illecite, aumentando il rischio di esposizione indiretta attraverso fornitori, partner e terze parti.

In questo contesto, anche il ransomware si trasformerà. Non sarà più solo uno strumento di blocco dei sistemi, ma una leva di estorsione articolata, capace di combinare furto di dati, pressione reputazionale

e sfruttamento di accessi persistenti. Il danno economico e operativo di un singolo incidente tenderà quindi ad aumentare, rendendo la resilienza e la capacità di recupero elementi centrali della strategia di sicurezza.

Il quadro sarà ulteriormente complicato dalle dinamiche geopolitiche. Gli Stati investiranno in modo crescente in capacità cyber basate su AI, integrandole in strategie di guerra ibrida e competizione economica. Anche le organizzazioni private, soprattutto nei settori strategici, potranno diventare bersagli indiretti di queste dinamiche, esponendosi a rischi che vanno oltre il perimetro tradizionale della criminalità informatica.

In risposta a questo scenario, le funzioni di sicurezza adotteranno in modo sempre più strutturale strumenti difensivi basati su AI. I Security Operations Center evolveranno verso modelli più automatizzati, in cui l'AI supporterà l'analisi comportamentale, la correlazione degli eventi e le decisioni operative. Tuttavia, questa evoluzione introdurrà nuove dipendenze dalla qualità dei dati, dalla trasparenza dei modelli e dalla loro corretta supervisione.

Diventerà quindi imprescindibile affrontare il tema della governance dell'AI. La mancanza di policy chiare, responsabilità definite e controlli sui sistemi intelligenti esporrà le organizzazioni non solo a rischi di sicurezza, ma anche a conseguenze normative, legali e reputazionali, in un contesto regolatorio sempre più stringente.

In conclusione, il 2026 segnerà il passaggio definitivo verso una cybersecurity intesa come leva strategica di governo dell'incertezza. Le organizzazioni che sapranno anticipare il rischio, governare l'AI e integrare la sicurezza nella strategia complessiva saranno quelle meglio posizionate per affrontare un contesto digitale sempre più complesso e interconnesso.

Il framework di sicurezza AI

Il framework PROMETHEUS AI di Maticmind è stato concepito per offrire un approccio concreto alla sicurezza e alla conformità dei sistemi di intelligenza artificiale. La sua struttura si basa su 10 moduli interconnessi che coprono l'intero ciclo di vita dell'AI, assicurando che i controlli tecnici e organizzativi non siano un ripensamento, ma parte integrante del processo di sviluppo.

Il framework introduce concetti avanzati come il Model Bill of Materials (MBOM), un inventario completo di tutti i componenti (dataset, modelli di base, librerie) utilizzati per costruire un modello di IA. Questo strumento è fondamentale per la tracciabilità e la sicurezza della supply chain, riducendo i rischi di compromissione e accelerando i processi di audit. Parallelamente, il Threat Modeling AI-specifico permette di identificare scenari di attacco unici per l'IA, come il poisoning dei dati o model evasion, prima che diventino vulnerabilità in produzione.

A livello di test e valutazione, PROMETHEUS AI non si limita ai test di sicurezza tradizionali, ma incorpora il Penetration Testing Linguistico (PT Linguistico), una pratica essenziale per individuare e mitigare le minacce specifiche dei modelli linguistici, come il prompt injection e il jailbreak. Questa attenzione al dettaglio si estende alla fase di Deployment & Protezione Runtime, dove l'uso di tecnologie come eBPF (extended Berkeley Packet Filter) consente di creare una "sandbox" per i modelli, monitorando e prevenendo comportamenti anomali in tempo reale, un approccio più efficace rispetto ai sistemi di DLP (Data Loss Prevention) standard che non riescono a intercettare attacchi di data exfiltration attraverso i modelli.

I benefici di questo framework sono significativi: si stima una riduzione dell'85% degli incidenti di sicurezza e un aumento del 30% nella velocità di rilascio in produzione, poiché i controlli di sicurezza vengono automatizzati e integrati nelle pipeline CI/CD. Inoltre, PROMETHEUS AI è progettato per garantire il 100% di conformità all'AI Act per i sistemi ad alto rischio, offrendo un percorso chiaro e documentabile verso la certificazione e la trasparenza. Questo non solo mitiga i rischi legali e reputazionali, ma costruisce anche la fiducia degli utenti e delle parti interessate.

Cyber Framework Maticmind

Il Cyber Framework Maticmind si allinea alla Direttiva NIS2 attraverso un modello integrato che anticipa minacce emergenti, riduce rischi con analisi proporzionate, promuove la sicurezza "by design", implementa tecnologie avanzate (EDR, SIEM, zero trust) e rafforza la resilienza tramite monitoraggio continuo e risposta rapida agli incidenti.

Questo approccio trasforma la compliance normativa in vantaggio strategico, garantendo continuità operativa, mitigazione degli impatti economici degli attacchi e una governance cyber dinamica.

Predittiva – Anticipare

Monitoraggio costante del panorama delle minacce tramite Cyber Threat Intelligence e tecniche OSINT, per identificare rischi emergenti prima che si manifestino.

Preventiva – Prevenire

Valutazione e gestione del rischio tecnologico, umano, organizzativo e compliance, per rafforzare la postura di sicurezza e ridurre la probabilità di incidenti.

Progettuale – Costruire Sicuro

Integrazione della sicurezza nei processi di design e architettura, secondo principi di "secure by design" e "security by resilience", per garantire sistemi resilienti.

Produttiva – Implementare Soluzioni

Adozione e gestione di tecnologie di difesa (firewall, EDR, IAM, SIEM) che trasformano le strategie in protezione operativa e misurabile.

Proattiva – Reagire e Migliorare

Monitoraggio continuo, risposta agli incidenti e analisi forense: la sicurezza come processo dinamico e adattivo, guidato dal miglioramento continuo.

Cinque centri di competenza, un'unica forza cyber

1

Centro strategico per l'analisi e gestione del rischio cyber: governance, compliance, modelli di maturità, supply chain security posture.

- Offensive Security
- Consulting Services (CisoasaService, PM,..)
- Governance, Risk e Compliance
- Social Engineering Simulation
- Risk Assessment
- Maturity Model checkup
- Supply Chain Security Posture
- Cyber Academy

2

Continuità operativa e sicurezza in tempo reale assicurata da presidio operativo che integra SOC H24, con le capacità tecniche del TAC e il controllo infrastrutturale del NOC. Garantisce un monitoraggio continuo, una risposta tempestiva agli incidenti e un supporto tecnico evoluto.

- **SOC (Security Operation Center)**
  - SOC H24
  - Cyber Threat Intelligence
  - Malware Analysis
- **SOS (SECURITY OPERATION SERVICES)**
  - Crisis Management
  - IncidentResponse & Digital Forensic Analysis
- **TAC (Technical Assistance Center)**
  - Gestione di ticket e supporto tecnico avanzato
  - Assistenza su soluzioni proprietarie e terze parti
  - Troubleshooting di sicurezza e infrastruttura
  - Supporto su incidenti tecnici e configurazioni complesse
- **NOC (Network Operation Center)**
  - Monitoraggio infrastrutturale H24
  - Supervisione performance di rete, cloud e sistemi
  - Gestione allarmi e segnalazioni
  - Coordinamento tecnico con il SOC in caso di anomalie

3

Esecuzione ed implementazione di progetti complessi di sicurezza: infrastrutture, cloud, OT/IoT, application security e integrazione di soluzioni cyber.

- Infrastructure & Network Security Management
- Cyber Resiliency e Data Protection
- Microsoft, Fortinet & CISCO Competence Center
- Cloud Security By Design
- Application Security SSDLC
- OT & IoT Security Management
- Security System Integration and Monitoring

4

Gestione e integrazione di soluzioni proprietarie e tecnologie di terze parti: piattaforme EDR, firewall, IAM, SIEM, DLP.

- Piattaforma Twin4Cyber
- Cyber Range
- WiseView - Risk Management Platform
- SIEM/SOAR
- EDR/XDR/NDR
- Identity & Access Management (IAM/PAM)
- NGFW
- Data Loss Prevention

5

Progetti speciali e sperimentazione: architetture avanzate, droni, AI, cyber-physical systems, prototipi e soluzioni non convenzionali.

- Architetture avanzate multi-layer e ibride
- Sistemi di comando e controllo per droni e UAV
- Integrazione AI in contesti cyber e sicurezza industriale
- Progetti su ambienti cyber-physical / smart infrastructure
- Soluzioni sperimentali
- Cyber testbed per ambienti OT e industriali
- Soluzioni cyber per difesa e protezione civile
- Soluzioni sperimentali edge & fog computing.

Disclaimer

Il presente report è stato elaborato utilizzando esclusivamente fonti di informazioni open source (OSINT) e dati pubblicamente accessibili. Le informazioni contenute in questo documento riflettono lo stato dell'arte al momento della sua redazione.

L'azienda che ha commissionato la creazione di questo report è esonerata da qualsiasi responsabilità riguardante:

- L'accuratezza, completezza o validità delle informazioni presentate
- Eventuali errori od omissioni nei dati analizzati
- Qualsiasi danno diretto o indiretto derivante dall'utilizzo delle informazioni contenute
- L'interpretazione dei dati e delle analisi presentate

Le conclusioni e le raccomandazioni fornite nel report sono basate sulle evidenze disponibili al momento della ricerca e non costituiscono garanzia di risultati futuri. Il panorama delle minacce cyber è in continua evoluzione, pertanto le informazioni potrebbero diventare obsolete rapidamente.

Questo documento è stato creato con l'unico scopo di fornire una panoramica generale sullo stato della cybersecurity non deve essere considerato come consulenza professionale o legale.

L'azienda che ha commissionato il report declina ogni responsabilità per:

- Decisioni operative o strategiche prese sulla base delle informazioni contenute
- Eventuali conseguenze derivanti dall'implementazione delle raccomandazioni suggerite
- L'uso improprio delle informazioni presentate

Si raccomanda vivamente ai lettori di consultare esperti del settore prima di intraprendere azioni basate sul contenuto di questo report.

MATICMIND
ZENITA GROUP

Bibliografia

- **ACI Worldwide** (2024). Prime Time for Real-Time Report.
- **Adversa AI** (2025). Top AI Security Incidents (2025 Edition). Disponibile presso: https://adversa.ai
- **Anthropic Research** (2025). Investigation into the GTG-1002 Campaign: AI as an Orchestrator of Cyber-Espionage.
- **Clusit – Associazione Italiana per la Sicurezza Informatica** (2025). Rapporto Clusit 2025 sulla sicurezza ICT in Italia.
- **Council on Foreign Relations (CFR)**. Cyber Operations Tracker. Disponibile presso: https://www.cfr.org/cyber-operations/
- **CrowdStrike** (2025). 2025 Global Threat Report. Disponibile presso: https://www.crowdstrike.com/en-us/global-threat-report/
- **ENISA – European Union Agency for Cybersecurity** (Ottobre 2025). ENISA Threat Landscape 2025. ISBN 978-92-9204-723-8, doi: 10.2824/1946374.
- **Global Cybersecurity Association (GCA)** (2025). Defending Against State-Sponsored Cyberattacks in 2025. Disponibile presso: https://gca.isa.org/blog/defending-against-state-sponsored-cyberattacks-in-2025
- **IDC – International Data Corporation** (2024). Worldwide Semi-annual Artificial Intelligence Tracker.
- **Juniper Research** (2023). Online Payment Fraud: Emerging Threats, Segment Analysis & Market Forecasts 2023-2027.
- **KELA Cyber Intelligence** (2025). 2025 AI Threat Report: How cybercriminals are weaponizing AI technology. A Guide to Understanding and Managing Emerging AI Cyber Threats.
- **Microsoft** (2024). Microsoft Digital Defense Report 2024.
- **Microsoft** (2025). Microsoft Digital Defense Report 2025 (MDDR).
- **Microsoft News** (2025, 28 gennaio). Governments Face Unprecedented Cyber Threats: AI Emerges as the Ultimate Defense to Cybercrime. Microsoft CEE Multi-Country News Center.
- **Microsoft Threat Intelligence** (2024). Staying ahead of threat actors using AI.
- **Telsy (TIM Group)** (2025). Cyber Security Report 2025.

MATICMIND

ZENITA GROUP

UNA CYBER ACIES

www.maticmind.it