L'Intelligenza Artificiale per la Cybersecurity: opportunità e sfide nella sicurezza digitale

di Franco Campitelli*

L'evoluzione dell'Intelligenza Artificiale sta modificando rapidamente il panorama della cybersicurezza. In questo saggio l'autore intende esaminare potenzialità e criticità del rapporto tra IA e *cybersecurity.* Se da un lato le minacce diventano sempre più sofisticate, dall'altro si stanno studiando strumenti in grado di contrastarle quali, ad esempio, l'analisi predittiva, l'identificazione di schemi di attacco in tempo reale e l'automatizzazione delle risposte ad incidenti di sicurezza. Ulteriori aspetti da considerare nell'utilizzo dell'IA sono la privacy, la protezione dei dati personali e i bias algoritmici. Infine, si analizzeranno idee per garantire i diritti della cittadinanza e minimizzare i rischi quali adozione di meccanismi di supervisione e tecniche di "*Differential Privacy*". In definitiva, il successo dell'IA nella cybersicurezza dipenderà dalla capacità di tutti gli *stakeholders* di assicurare la protezione dei sistemi con un'attenzione particolare alla tutela della privacy e dei diritti fondamentali.

Parole chiave: intelligenza artificiale; cybersecurity; minacce; privacy; bias algoritmici.

Artificial intelligence for cybersecurity: opportunities and challenges in digital security

The rapid evolution of artificial intelligence is changing the cybersecurity landscape. In this essay, the author examines the potential and critical issues surrounding the relationship between AI and cybersecurity. As threats become more sophisticated, tools are being developed to counter them, including predictive analysis, real-time identification of attack patterns, and automated responses to security incidents. Other considerations when using AI include privacy, personal data protection and algorithmic bias. Finally, the essay will analyze ideas for guaranteeing citizens' rights and minimizing risks, such as the adoption of oversight mechanisms and 'differential privacy' techniques. Ultimately, the success of AI in cybersecurity depends on all stakeholders' ability to protect systems, with a particular focus on privacy and fundamental rights.

Keywords: artificial intelligence; cybersecurity; threats; privacy; algorithmic bias.

DOI: 10.5281/zenodo.18435761

* Università degli Studi di Teramo, fcampitelli@unite.it

Franco Campitelli

Introduzione

Oggi viviamo in un'epoca di enorme trasformazione digitale in cui i sistemi informatici, i dati e le persone sono sempre più interconnessi tra loro. Se da un lato questa interconnessione ha reso più rapide le comunicazioni e ha migliorato significativamente l'efficienza sul lavoro, la stessa ha fornito un enorme valore ai "dati digitali" che transitano in rete. Quando si acquista online, si fa una ricerca, si naviga all'interno dei siti tutti i nostri spostamenti online sono tracciati dai cosiddetti "*cookies*". È evidente l'importanza che questo piccolo file riveste per un'azienda; infatti, in tempo reale è possibile studiare il comportamento online di milioni di persone e valutarne i gusti e le tendenze da utilizzare in campagne pubblicitarie mirate. I *social* in questo senso sono diventati dei "forzieri" di informazioni tanto che Floridi ha coniato il cosiddetto termine "infosfera"[2] per descrivere il mondo digitale in cui siamo immersi. Valutata l'importanza che il dato oggi riveste nel mondo digitale diventa naturale per i cosiddetti "*black hat hacker*" cercare di impossessarsene per rivenderli al miglior offerente. In questo scenario entra in maniera dirompente l'Intelligenza Artificiale che ha permesso di velocizzare molte operazioni comprese quelle di *coding* sia benevole che malevole, pertanto, le difese tradizionali utilizzate fino a qualche anno fa, diventano inadeguate. Ulteriori aspetti da considerare sono, inoltre, la tutela della privacy e la riduzione dei bias algoritmici. Il presente contributo si propone di analizzare il rapporto tra intelligenza artificiale e cybersecurity da più punti di vista quali le opportunità di potenziare la sicurezza digitale con l'IA esaminando le minacce, valutando le difese digitali, analizzando le sfide, i rischi e le implicazioni etico-sociali. Infine, si proporranno strategie di mitigazione, governance e prospettive future.

1. Il Panorama Attuale: Minacce Cibernetiche e il Ruolo Emergente dell'IA

1.1 Evoluzione delle minacce cibernetiche

In questo capitolo si esplorerà il panorama attuale delle minacce informatiche sulla base delle informazioni fornite da report recenti di organizzazioni leader nel settore. Conoscere le minacce più diffuse, chi le guida e le tattiche

[2] L'infosfera è la globalità dello spazio delle informazioni: un ambiente che comprende tutti i flussi informativi, sia digitali che analogici, che attraversano la nostra società (Floridi, 2020).

utilizzate diventa fondamentale per lo sviluppo di strategie di difesa efficaci. Gli attori esterni *(*i c.d. *black hat hacker)* o loro organizzazioni (*Anonymous)* sono spesso motivati da fattori quali ideologia, spionaggio o finanziari. Secondo le analisi condotte da Verizon "2025 Data Breach Investigations Report" la principale categoria di attacco è l'intrusione di sistema (*System Intrusion*) seguita dal *Social Engineering, Basic Web Application Attacks* (BWAA) oltre all'abuso di privilegi (*Privilege Misuse*). Un ulteriore elemento da non sottovalutare è il coinvolgimento di terze parti nelle violazioni come fornitori esterni di servizi e gestori di dati che fanno parte della cosiddetta *supply-chain*. Un esempio recente nel 2023 ha coinvolto il software «MOVEit»[3]. I malintenzionati hanno sfruttato vulnerabilità multiple e attraverso l'esecuzione di codice remoto hanno compromesso oltre 2600 organizzazioni impossessandosi di dati personali sensibili quali indirizzi del personale, documenti di identità e numeri di carte di credito. Questo incidente ha ulteriormente evidenziato come, in una società connessa, un punto debole in un software di terze parti possa compromettere dati sensibili e causare dannose interruzioni operative compromettendo la continuità aziendale delle aziende coinvolte. Un ulteriore fattore di rischio è dato dal "fattore umano". Un utente inesperto o "distratto" potrebbe cadere nella trappola del *social engineering*. Si citano due esempi significativi: il primo caso avviene nel dipartimento del lavoro degli Stati Uniti dove è stato utilizzato un sistema di phishing con domini falsificati (*spoofing*) per rubare credenziali di accesso a Office365, il secondo caso ha preso di mira un'azienda produttrice di aeromobili che è stata di vittima di una truffa di *Business E-mail Compromise* (BEC). È stato violato l'account di posta elettronica dell'amministratore delegato per inviare una richiesta "urgente" di trasferimento di fondi causando una perdita di molti milioni di euro. Un ulteriore attacco che non prende di mira le credenziali, ma può essere causa di disagi è il *Distributed Denial of Service* (DDoS); spesso è attuato contro l'amministrazione pubblica, i trasporti, gli ospedali e il settore bancario.

1.2 Limiti degli approcci tradizionali alla cybersecurity

L'evoluzione dell'Intelligenza Artificiale e del Machine Learning applicata alla cybersicurezza sta rendendo rapidamente obsoleti gli approcci

[3] Software utilizzato per il trasferimento sicuro di file tra organizzazioni.

tradizionali. Le tecniche basate sulle firme (*signature-based*) che utilizzano la comparazione del traffico di rete con un database di regole e firme non possono rilevare attacchi nuovi o "*zero-day"* e hanno bisogno di frequenti aggiornamenti manuali. La dinamicità e la variabilità delle minacce, di conseguenza, dovranno andare oltre il riconoscimento degli schemi conosciuti applicando l'IA e il ML attraverso le tecniche basate sulle anomalie (*anomaly-based*). Se da un lato questa tecnica permette di rilevare minacce sconosciute e non richiede aggiornamenti continui delle firme, dall'altro potrebbe generare un numero elevato di "falsi positivi". Uno studio di Jada e Mayayise ha evidenziato che «gli approcci basati sull'AI superano i metodi non basati sull'AI in termini di efficacia e precisione per il rilevamento delle intrusioni» (2024: 5).

1.3 Introduzione dell'AI come risposta

Una definizione di Intelligenza Artificiale, ripresa da Russel e Norvig, è «lo studio degli agenti che ricevono percezioni dall'ambiente ed eseguono azioni» (2016: VIII). Nel cosiddetto "modello standard" un agente "razionale" agisce in maniera tale da "massimizzare il valore atteso di una misura di prestazione, data la sequenza percettiva fino a quel momento". Lo sviluppo dell'IA negli ultimi anni ha seguito due approcci differenti. Dapprima, con l'introduzione di ChatGPT nel 2022, si è assistito allo sviluppo di modelli *Large Language Model* (LLM), chatbot specializzati nell'elaborazione e generazione del linguaggio naturale che hanno mostrato una buona capacità di comprensione del testo e nella produzione di risposte coerenti. Gli LLM, però, non possono agire autonomamente ed è per questo motivo che si stanno creando Agenti AI per compiti più specializzati quali assistenza clienti, supporto interno alle aziende, sanità ecc. Questa automazione spinta comporterà una valutazione sui rischi e sulle implicazioni etico-sociali che verranno sviluppate nei capitoli successivi. Nel campo della cyber-security l'utilizzo dell'IA rappresenta uno strumento efficace per migliorare le misure di sicurezza.

Franco Campitelli

2. Opportunità dell'Intelligenza Artificiale per la Difesa Digitale

2.1. Rilevamento e Prevenzione Avanzata

Come già dimostrato nel capitolo precedente, negli ultimi anni i metodi di difesa tradizionali basati sulle firme applicati contro le minacce informatiche si stanno rilevando insufficienti. In questo capitolo si esamineranno alcune applicazioni pratiche dell'applicazione dell'IA e del ML nella cybersicurezza. Due pratiche principali sono normalmente applicate nell'ambito della sicurezza: l'analisi comportamentale e il rilevamento delle anomalie. L'analisi comportamentale è una tecnica che consiste nell'osservare, raccogliere e analizzare il modo di agire degli utenti per individuare comportamenti anomali e/o minacce. Si procede dapprima con la raccolta dei dati (clic, movimenti del mouse, utilizzo di applicazioni), si crea un profilo comportamentale "normale" utilizzando metodi statistici e intelligenza artificiale. Il rilevamento delle anomalie, invece, si basa sull'identificazione di elementi o eventi che si discostano significativamente dal modello normale o atteso all'interno di un determinato set di dati (Katiyar *et al*., 2024). In cybersecurity, questa tecnica può essere utilizzata per identificare attività sospette o dannose all'interno di una rete o di un sistema. L'IA può essere utilizzata anche in fase preventiva per esaminare i sistemi ricercando vulnerabilità note, raccomandando modifiche a firewall, sistemi di prevenzione delle intrusioni e altri controlli di sicurezza in base alle evoluzioni delle minacce (Roshanaei *et al*., 2024).

2.2. Automazione e risposta

Una rapida risposta agli attacchi informatici può essere implementata con l'utilizzo di AI e ML. Queste nuove tecnologie permettono l'analisi di grandi quantità di dati in tempo reale provenienti da più fonti quali traffico di rete, registri di sistema e attività degli utenti. A tal proposito sono stati sviluppati i Security Information and Event Management (SIEM). Un esempio concreto di SIEM è l'IBM QRadar che dopo aver raccolto e normalizzato i dati, analizzato gli eventi e rilevate le eventuali minacce utilizza la tecnica di analisi *User and Entity Behaviour Analytics (UEBA)* e tramite il *Security Orchestration, Automation and Response (SOAR)* automatizza i processi di risposta

agli incidenti permettendo di condurre anche analisi forensi sugli incidenti. In questo modo le aziende non sono più costrette a lunghe analisi e ricerche condotte da esperti grazie all'automazione dei processi che permette di ampliare l'efficacia operativa e migliorare la sicurezza (Mandru, 2022). I moderni sistemi, potenziati dai LLM e dagli agenti IA, riescono ad interpretare anche dati non strutturati adattandosi a minacce sconosciute e automatizzando flussi di lavoro complessi (Ismail *et al*., 2025). Di conseguenza l'analisi in tempo reale continua da set di dati estesi unita alla capacità dell'IA risulta notevolmente più efficiente di semplici sistemi di correlazione basati su regole predefinite rendendo le misure di sicurezza più dinamiche e proattive (Mohamed, 2025).

3. Sfide, Rischi e Implicazioni Etico-Sociali dell'IA nella Cybersecurity

3.1. Sfide e rischi dell'IA

La convergenza tra intelligenza artificiale e sicurezza informatica sta provocando un cambiamento di prospettiva. Se da un lato rappresenta un grande potenziale da sfruttare, dall'altro potrebbe essere associata anche a complessità interne. L'integrazione dell'IA nella sicurezza informatica, di conseguenza, introduce una serie di sfide, rischi e implicazioni etico-sociali che meritano un'attenta riflessione (Achuthan *et al*., 2024). Una delle sfide principali risiede nel fatto che gli algoritmi di IA e le tecniche possono essere utilizzati sia a fini difensivi che per attività dannose (Titus, Russell, 2023). Gli strumenti basati sull'IA, come approfondito in precedenza, possono automatizzare il rilevamento e la risposta alle minacce, ma possono anche essere sfruttati dai criminali informatici per sviluppare malware sofisticati, campagne di phishing e attacchi di ingegneria sociale (Morla, 2019). Di conseguenza, i rischi legati a tutto ciò che sottende ai dati, comprese le questioni relative alla condivisione, ai bias e al cosiddetto "data poisoning", rappresentano importanti preoccupazioni poiché l'apprendimento automatico dipende da enormi quantità di dati generati dall'uomo per l'addestramento. Un'altra sfida significativa riguarda la natura degli algoritmi di IA. Questi spesso rappresentano una "black box" rendendo difficile, se non impossibile, comprendere il percorso seguito per arrivare alla decisione. Questa mancanza di trasparenza potrebbe minare la fiducia nei sistemi di sicurezza basati

sull'IA. Una ulteriore preoccupazione potrebbe sorgere quando si parla di responsabilità: in particolare quando i sistemi di IA prendono decisioni che hanno conseguenze significative oppure per l'imprevedibilità degli sviluppi della tecnologia. Diventa quindi necessario, per limitare questi rischi, definire strategie precise quali creazione di robusti protocolli di sicurezza, approcci di sicurezza multilivello e l'uso della tecnologia IA (Xu *et al*., 2024).

3.2. Implicazioni etiche e sociali nell'uso dell'IA nella sicurezza informatica

L'implementazione dell'IA nella sicurezza informatica solleva diverse considerazioni etiche e sociali che devono essere affrontate in modo fattivo. Gli algoritmi di IA potrebbero essere soggetti ai cosiddetti "bias algoritmici" risultando in output discriminatori. Secondo Choung (2023), se i dati utilizzati per addestrare i sistemi di IA riflettessero i pregiudizi sociali esistenti, questi potrebbero essere addirittura amplificati. Un esempio è il programma COMPAS utilizzato negli Stati Uniti. Nello studio di Angwin *et al*. (2016) è stato messo in discussione l'utilizzo della giustizia predittiva a causa della presenza di bias all'interno dell'algoritmo. Questi sistemi, se non correttamente utilizzati, potrebbero violare i diritti umani di determinati gruppi demografici e sollevare preoccupazioni in merito alla violazione della privacy.

3.3. Come è possibile affrontare le sfide poste dall'uso dell'IA?

Molti sono i rischi e le implicazioni etico-sociali correlati all'utilizzo dell'IA nella sicurezza informatica. A parere di chi scrive diventa essenziale un approccio collaborativo e multidisciplinare. Sarebbe auspicabile la collaborazione di esperti di vari settori, tra cui informatica, ingegneria, diritto, etica e scienze sociali, al fine di studiare e sviluppare soluzioni complete che affrontino gli aspetti tecnici, giuridici e sociali della sicurezza basata sull'IA. Diventa inoltre necessario un dialogo e un impegno costanti tra responsabili politici, società leader del settore, ricercatori e persone comuni, con l'obiettivo di comprendere le opportunità e risolvere le sfide presentate dall'uso dell'IA. A livelli più alti è fondamentale la cooperazione internazionale per affrontare le questioni relative alla governance dei dati, alla criminalità informatica e alla sovranità digitale. In definitiva solo affrontando in modo dinamico queste sfide è possibile migliorare la sicurezza informatica

mitigandone i rischi e garantendone un uso responsabile ed etico (Jada, Mayayise, 2023).

4. Strategie di Mitigazione, Governance e Prospettive Future

4.1. Strategie di mitigazione

Per diminuire e ridurre i rischi dovuti all'uso dell'IA nel contesto della cybersecurity è necessario utilizzare un approccio diversificato integrando soluzioni tecnologiche all'avanguardia unite a solide politiche di governance (Schmitt, Koutroumpis, 2025). Bisogna attuare strategie di threat hunting basate su algoritmi di machine learning che possono identificare e neutralizzare minacce note e sconosciute prima di danneggiare significativamente i sistemi informatici (Xu *et al*., 2024). Per mantenere un sistema di protezione nel tempo è possibile implementare il cosiddetto "continuous learning" che prevede l'aggiornamento costante delle competenze, delle conoscenze e delle pratiche dei professionisti e delle organizzazioni per fronteggiare l'evoluzione rapida delle minacce informatiche. Anche la protezione della privacy dei dati utilizzati per l'addestramento dei modelli di IA rappresenta un aspetto fondamentale che richiede l'implementazione di tecniche avanzate di anonimizzazione e crittografia. In definitiva oltre alle tecniche esaminate in precedenza, si può affermare che la qualità e la differenziazione dei dati utilizzati per l'addestramento migliorano notevolmente l'efficacia delle misure di sicurezza basate sull'IA (Alevizos, Dekker, 2024).

4.2. Governance dell'IA applicata alla cybersecurity

La governance dell'IA applicata alla cybersecurity deve essere guidata da principi etici e legali ben definiti. Affinché i cittadini ripongano fiducia nell'IA, è necessario che gli algoritmi siano trasparenti e che le decisioni prese siano comprensibili e giustificabili (Achuthan *et al*., 2024). I sistemi di IA devono essere progettati in modo da evitare bias e discriminazioni, assicurando un trattamento equo per tutti i cittadini; ciò può essere messo in pratica mediante l'implementazione di procedure di controllo e supervisione umana (Cath, 2018). È essenziale, inoltre, che i sistemi siano conformi a

normative internazionali (GDPR e AI ACT in Europa) per garantire la protezione dei dati personali (Bharati, 2024).

4.3. Prospettive future

Per il futuro si prevede che l'Intelligenza Artificiale si sviluppi in maniera molto rapida e di conseguenza si dovranno affrontare minacce sempre più complesse. Un'idea per migliorare i sistemi di sicurezza potrebbe essere quella di integrare l'IA con tecnologie emergenti quali cloud computing, IoT e blockchain. Un ulteriore passo in avanti si potrà fare sviluppando i sistemi con la cosiddetta eXplainable AI (XAI) che si basa su principi di trasparenza, interpretabilità, controllabilità e validità, con l'obiettivo di rendere i modelli di intelligenza artificiale comprensibili e affidabili per gli utenti, aumentando la fiducia e facilitando l'adozione in contesti critici. Dal punto di vista legislativo bisognerà regolamentare sempre meglio l'IA trovando un giusto bilanciamento tra innovazione e privacy.

Conclusioni

In conclusione, l'integrazione dell'intelligenza artificiale nel campo della cybersecurity rappresenta un'evoluzione cruciale nel panorama della sicurezza digitale contemporanea segnato da minacce cibernetiche sempre più sofisticate e pervasive (Bonfanti, 2022). L'IA diventa uno strumento di difesa avanzato in continua evoluzione capace di migliorare le strategie di protezione e di anticipare le mosse degli attaccanti con una velocità e una precisione senza precedenti (Ofusori *et al*., 2024). L'abilità di apprendere e adattarsi continuamente, tipica degli algoritmi di machine learning, consente ai sistemi di sicurezza basati sull'IA di evolvere in risposta alle nuove minacce, superando i limiti degli approcci tradizionali, basati su regole statiche e firme predefinite. Tuttavia, l'adozione dell'IA nella cybersecurity non è priva di sfide e implicazioni complesse. La mancanza di trasparenza nei processi decisionali degli algoritmi di IA, la vulnerabilità degli algoritmi di IA agli attacchi adversarial, rappresentano una minaccia concreta alla loro efficacia. È fondamentale considerare le implicazioni etiche e legali dell'utilizzo dell'IA nella cybersecurity, garantendo che le tecnologie siano impiegate in modo responsabile e nel rispetto dei diritti fondamentali, come la privacy e la protezione dei dati personali. In definitiva, a parere di chi scrive, per il futuro è

necessario tendere ad un percorso in cui siano bilanciati sicurezza e rispetto dei diritti fondamentali.

Riferimenti bibliografici

Achuthan K., Ramanathan S., Srinivas S., Raman R. (2024). Advancing cybersecurity and privacy with artificial intelligence: current trends and future research directions. *Frontiers in Big Data*, 7. DOI: 10.3389/fdata.2024.1497535.

Alevizos L., Dekker M. (2024). Towards an AI-enhanced cyber threat intelligence processing pipeline. *arXiv* (Cornell University). DOI: 10.3390/electronics13112021.

Angwin J., Larson J., Mattu S., Kirchner L. (2016). Machine bias. *ProPublica*.

Bharati R. (2024). The right to privacy in the age of artificial intelligence: challenges and legal frameworks. *SSRN*. https://ssrn.com/abstract=4908340 (consultato il …).

Bonfanti M. (2022). Artificial intelligence and the offense-defense balance in cybersecurity. In: *Artificial Intelligence and International Security*. DOI: 10.4324/9781003110224-6.

Cath C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A*, 376: 20180080. DOI: 10.1098/rsta.2018.0080.

Choung H., David P., Seberger J.S. (2023). A multilevel framework for AI governance. *arXiv* (Cornell University). DOI: 10.48550/arXiv.2307.03198.

Floridi L. (2020). *Pensare l'infosfera*. Milano: Raffaello Cortina Editore.

Ismail I., Kurnia R., Brata Z.A., Nelistiani G.A., Heo S., Kim H. (2025). Toward robust security orchestration and automated response in security operations centers with a hyperautomation approach using agentic artificial intelligence. *Information*, 16(5): 365. DOI: 10.3390/info16050365.

Jada I., Mayayise T.O. (2024). The impact of artificial intelligence on organisational cybersecurity: an outcome of a systematic literature review. *Data and Information Management*, 8(2). DOI: 10.1016/j.dim.2023.100063.

Katiyar N., Tripathi S., Kumar P., Verma S., Kumar S.A., Saxena S. (2024). AI and cybersecurity: enhancing threat detection and response with machine learning. *Educational Administration: Theory and Practice*, 30(4): 6273-6282. DOI: 10.53555/kuey.v30i4.2377.

Mandru S. (2022). How AI can improve identity verification and access control processes. *Journal of Artificial Intelligence and Cloud Computing*, 1. DOI: 10.47363/jaicc/2022(1)e101.

Mohamed N. (2025). Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms. *Knowledge and Information Systems*. DOI: 10.1007/s10115-025-02429-y.

Morla R. (2019). Ten AI stepping stones for cybersecurity. *arXiv* (Cornell University). DOI: 10.48550/arXiv.1912.06817.

Ofusori L., Bokaba T., Mhlongo S. (2024). Artificial intelligence in cybersecurity: a comprehensive review and future direction. *Applied Artificial Intelligence*. DOI: 10.1080/08839514.2024.2439609.

Roshanaei M., Khan M.R., Sylvester N.N. (2024). Enhancing cybersecurity through AI and ML: strategies, challenges, and future directions. *Journal of Information Security*, 15: 320-339. DOI: 10.4236/jis.2024.153019.

Schmitt M., Koutroumpis P. (2025). Cyber shadows: neutralizing security threats with AI and targeted policy measures. *IEEE Transactions on Artificial Intelligence*. DOI: 10.1109/TAI.2025.3527398.

Titus A.J., Russell A.H. (2023). The promise and peril of artificial intelligence: violet teaming offers a balanced path forward. *arXiv* (Cornell University). DOI: 10.48550/arXiv.2308.14253.

Xu H., Li Y., Balogun O., Wu S., Wang Y., Cai Z. (2024). Security risks and concerns of generative AI in the IoT. *arXiv* (Cornell University). DOI: 10.48550/arXiv.2404.00139.