International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Jan-Feb-2026, ISSN (Online): 2395-566X

The Dual Role of Artificial Intelligence in Cyber Security: From Automated Defense to Adversarial Exploitation

Sachin Kumar
Computer Science & Engineering Dr. A.P.J. Abdul Kalam Technical University, Lucknow,
Jahangirabad Educational Trust Group of Institutions, Barabanki. UP.

Abstract- — The Dual Role of Artificial Intelligence in Cyber Security: From Automated Defense to Adversarial Exploitation Abstract The rapid integration of Artificial Intelligence (AI) into the digital landscape has fundamentally transformed the field of cyber security. This paper examines the bidirectional impact of AI: its role as a powerful defensive mechanism capable of real-time threat detection and response, and its emergence as a sophisticated tool for adversarial exploitation. By analyzing Machine Learning (ML) models in intrusion detection, the rise of "Agentic" autonomous security systems, and the threats posed by adversarial ML and deepfakes, this study proposes a framework for AI-resilient security operations. The research concludes that while AI significantly enhances defensive capabilities, it also necessitates a new era of proactive, adaptive security strategies to counter AI-driven threats.

Keywords: Artificial Intelligence (AI), Cyber Security, Machine Learning (ML), Agentic Systems, Adversarial AI, Deepfake Detection, Zero-Trust Identity.

I. INTRODUCTION

Traditional cyber security relies heavily on signature-based detection and manual intervention—methods that are increasingly inadequate against the volume and speed of modern cyber-attacks. The introduction of AI and Machine Learning has shifted the paradigm from reactive defense to proactive prediction. However, as defenders adopt AI to secure networks, attackers are simultaneously leveraging the same technologies to automate reconnaissance, bypass authentication, and create polymorphic malware. +2

II. AI AS A DEFENSIVE SHIELD

AI systems are uniquely suited for cyber security due to their ability to process massive datasets and identify patterns that are imperceptible to human analysts.

- Anomaly Detection and UEBA Machine Learning algorithms, particularly unsupervised learning models, are utilized for User and Entity Behavior Analytics (UEBA). By establishing a "baseline" of normal network behavior, AI can instantly flag deviations—such as a user accessing sensitive database records at an unusual hour—that may indicate a compromised account or an insider threat. +1
- Automated Incident Response (SOAR) Security Orchestration, Automation, and Response (SOAR) platforms now integrate AI to execute "Playbooks" without human intervention. For instance, if a system detects a high-confidence ransomware signature, the AI agent can autonomously isolate the infected endpoint from the network, preventing lateral movement in milliseconds.

III. THE RISE OF AGENTIC SECURITY SYSTEMS

A significant trend in 2026 is the transition from static AI models to Agentic Systems. These are autonomous agents capable of "reasoning" through a security incident.
Self-Healing Networks: These systems can identify a misconfiguration or a vulnerability and automatically generate and apply a patch.

Autonomous Threat Hunting: Instead of waiting for an alert, agentic AI proactively "hunts" through logs and network traffic to find dormant threats or "Low-and-Slow" attacks that bypass traditional thresholds.

IV. THE ADVERSARIAL FRONTIER

AI-Powered Attacks The "democratization" of AI has provided threat actors with powerful tools to enhance their offensive capabilities.

- Adversarial Machine Learning Attackers can target the security AI itself through Evasion Attacks (feeding the model subtly altered data to make it misclassify a threat as "clean") or Data Poisoning (corrupting the training data so the model learns to ignore specific malicious patterns).

- Enhanced Social Engineering Generative AI and Large Language Models (LLMs) enable attackers to create highly convincing, error-free phishing emails in multiple languages. Furthermore, Deepfake technology—utilizing AI to spoof audio or video—is being used in Business Email Compromise (BEC) attacks to impersonate executives and authorize fraudulent wire transfers. +1

V. SUSTAINABLE AND ETHICAL AI IN SECURITY

As AI models grow in complexity, their computational and environmental costs increase. This paper advocates for:
GreenOps for Security: Optimizing AI training cycles to reduce the carbon footprint of security operations.

Explainable AI (XAI): Ensuring that security AI provides transparent reasoning for its decisions, allowing human analysts to trust and verify autonomous actions.

VI. CONCLUSION

The intersection of AI and Cyber Security represents a permanent "Arms Race." While AI provides the only feasible path to securing the vast, high-speed networks of the future, it also introduces systemic risks that must be managed. The path forward lies in a hybrid model: leveraging the speed and scale of AI agents while maintaining human oversight for ethical judgment and strategic decision-making.

REFERENCES

1. Buczak, A. L., & Guven, E. (2016). "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection." IEEE Communications Surveys & Tutorials.
2. Goodfellow, I. J., et al. (2014). "Explaining and Harnessing Adversarial Examples." arXiv preprint.
3. National Institute of Standards and Technology (NIST). (2024). Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations.
4. Luccioni, A. S. (2023). "The Environmental Impact of Large Scale AI Models." Journal of Green Computing