



REVIEW ON ARTIFICIAL INTELLIGENCE IN CYBERSECURITY

DR. J. JOSELIN¹, DEV ANTHONY. P², SHRIMIKA. SK³

¹Associate Professor Department of Computer Applications/ Sri Krishna Arts and Science College / Coimbatore / Tamil Nadu

² Department of Computer Applications/ Sri Krishna Arts and Science College /Coimbatore / Tamil Nadu

³ Department of Computer Applications/ Sri Krishna Arts and Science College / Coimbatore / Tamil Nadu

ABSTRACT

As digital technologies continue to grow, cyber threats are becoming more frequent and complex, making traditional security methods less effective. Artificial Intelligence (AI) and Machine Learning (ML) have emerged as powerful technologies that help organizations detect, prevent, and respond to cyberattacks more efficiently. This review paper discusses the role of AI and ML in modern cybersecurity, covering techniques such as machine learning, deep learning, Large Language Models (LLMs), Explainable AI (XAI), and metaheuristic optimization algorithms. It explores their applications in intrusion detection, malware analysis, phishing detection, threat intelligence, and anomaly detection across cloud computing, IoT, and cyber-physical systems. The paper also highlights the growing use of Generative AI, which improves security capabilities but can also be misused to create advanced cyberattacks such as phishing and malware. In addition, it examines emerging technologies such as federated learning for privacy-preserving security and quantum-enhanced AI for future cyber defense. Finally, the

review emphasizes the importance of combining AI with human expertise to build transparent, reliable, and self-healing cybersecurity systems that can effectively protect modern digital environments.[2]

KEYWORD: Artificial intelligence, Cybersecurity, Machine learning, Intrusion detection, Adversarial ml, Llms, Explainable ai, Federated learning, Generative ai, 8569 Cyber defence

I. INTRODUCTION

In today's digital world, technology plays a vital role in daily life, with billions of smart devices and internet users generating vast amounts of data. While this growth has improved communication, business operations, and access to information, it has also increased opportunities for cybercriminals to exploit vulnerabilities.[5]

Cyber threats such as ransomware, malware, phishing attacks, and Advanced Persistent Threats (APTs) have become more sophisticated, making traditional security solutions less effective. Major

incidents like the WannaCry ransomware attack and the Yahoo data breach highlight the serious impact of cybersecurity failures.[9]

To address these challenges, organizations are increasingly adopting Artificial Intelligence (AI) and Machine Learning (ML). Unlike traditional rule-based systems, AI and ML can analyze large volumes of data, detect unusual patterns, and identify threats in real time. These technologies improve threat detection, automate security processes, and help organizations respond more quickly to cyberattacks.[2]

However, AI also presents new risks. Cybercriminals can exploit AI through techniques such as Adversarial Machine Learning, while AI-generated content can make phishing attacks and malware more convincing and effective.[4]

Despite these challenges, AI and ML remain valuable tools for enhancing cybersecurity. Emerging approaches such as Deep Learning, Federated Learning, and Explainable AI (XAI) are helping create more adaptive, secure, and trustworthy defense systems. This review explores the applications, benefits, challenges, and future potential of AI and ML in strengthening modern cybersecurity.[4]



Fig 1: AI-Driven Cybersecurity Architecture

II. Taxonomy of Machine Learning and Deep Learning in Cybersecurity

Machine Learning (ML) and Deep Learning (DL) play a key role in modern cybersecurity by improving threat detection, anomaly identification, and automated response. These techniques are mainly categorized into shallow learning paradigms, deep learning architectures, and metaheuristic algorithms.

A. Shallow Learning Paradigms

Shallow learning includes supervised, unsupervised, semi-supervised, ensemble, and reinforcement learning. These methods are used for malware detection, intrusion detection, anomaly detection, and adaptive cyber defense. Ensemble learning improves prediction accuracy by combining multiple models, while reinforcement learning enables intelligent and dynamic security responses.[1]

B. Deep Learning Architectures

Deep learning models such as CNNs, RNNs, LSTMs, DBNs, GANs, VAEs, and Large Language Models (LLMs) are used to analyze complex cybersecurity data. They help detect malware, identify attack patterns, perform anomaly detection, and support threat intelligence, log analysis, and automated security operations.[1]

C. Metaheuristic Algorithms

Metaheuristic algorithms improve the efficiency of machine learning models by optimizing feature selection and detection performance. Techniques such as Genetic Algorithms, Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Grey Wolf Optimizer, Simulated Annealing, and TLBO are widely used to strengthen intrusion detection and overall cybersecurity performance.[5]

III. STRATEGIC APPLICATION DOMAINS OF AI AND MACHINE LEARNING IN CYBERSECURITY

Artificial Intelligence (AI) and Machine Learning (ML) have become essential technologies in

modern cybersecurity by providing intelligent and automated solutions for detecting and preventing cyber threats. They are widely used in intrusion detection, malware and ransomware analysis, phishing detection, biometric authentication, IoT security, threat intelligence, and vulnerability management.

AI-based intrusion detection systems analyze network traffic and system data to identify attacks such as malware, insider threats, and DDoS attacks more effectively than traditional methods. Deep learning models help detect new malware variants by analyzing their behavior and code patterns. In IoT and smart environments, AI enables real-time anomaly detection to improve the security of connected devices.

Machine learning also enhances phishing detection by analyzing emails, URLs, and user behavior to identify fraudulent activities. AI-powered biometric systems strengthen authentication using facial recognition, fingerprints, and behavioral patterns. In Security Operations Centers (SOCs), AI automates threat intelligence, incident analysis, and response, reducing the time required to handle cyber incidents. Additionally, AI assists in vulnerability management by identifying software weaknesses, prioritizing risks, and supporting automated patch deployment. Overall, AI and ML help organizations build faster, smarter, and more resilient cybersecurity systems.[6]

IV. LARGE LANGUAGE MODELS (LLMs) AND GENERATIVE AI IN CYBER DEFENCE

Large Language Models (LLMs) and Generative AI have revolutionized cybersecurity by enhancing threat detection, automation, and decision-making capabilities. In defensive applications, they help identify software vulnerabilities, analyze code for security flaws, generate secure patches, support compliance monitoring, and provide understandable

explanations for complex AI-driven security decisions. However, these technologies also introduce new risks, as attackers can use them to automate reconnaissance, generate exploits, create adaptive malware, conduct prompt injection attacks, and launch highly convincing phishing campaigns. To address these evolving challenges, cybersecurity is increasingly adopting innovative approaches such as multi-agent AI systems that collaborate on threat analysis and incident response, psycholinguistic profiling of threat actors for improved attribution, and federated learning frameworks that enable organizations to share intelligence and train security models collaboratively while preserving data privacy.[3]

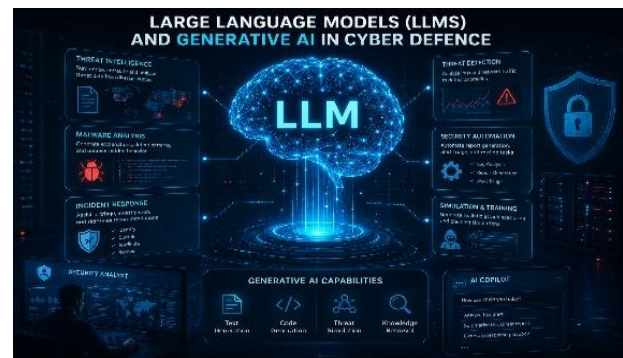


Fig 2 : LLMs in Cybersecurity

V. EXPLAINABLE AI (XAI) AND HUMAN-CENTRIC SECURITY

Explainable Artificial Intelligence (XAI) is important in cybersecurity because many AI and deep learning models function as black boxes, making their decisions difficult to understand. XAI improves transparency by providing clear explanations for threat detection and security decisions, helping analysts build trust in AI systems. Since AI cannot fully replace human expertise, Human-in-the-Loop approaches combine automated threat detection with human review and feedback. Frameworks such as Technology–Human–Task–Context (THTC) promote effective collaboration between AI and security professionals. However, explainability

can also introduce risks, such as information leakage and manipulation of explanations by attackers. Overall, XAI helps create more transparent, trustworthy, and human-centered cybersecurity systems while supporting better decision-making and regulatory compliance.[8]

Comparison of Explainable AI (XAI) and Human-Centric Security

Aspect	Explainable AI (XAI)	Human-Centric Security
Meaning	AI systems that provide clear and understandable explanations for their decisions.	Security approach that puts human needs, expertise, and decision-making at the center of the process.
Primary Focus	Making AI decisions transparent and interpretable.	Enhancing security by involving humans in the loop and using their judgment.
Decision Process	AI explains how and why a prediction or decision is made.	Humans analyze, validate and take final decisions with AI support.
Transparency	High - Provides reasons and logic behind AI outputs.	Ensures clarity by combining AI insights with human understanding.
Trust	Increases trust by making AI behavior understandable.	Builds trust through human responsibility and accountability.
Role of Humans	Humans understand AI explanations and supervise the system.	Humans are actively involved in monitoring, evaluating and responding to threats.
Advantages	Improved interpretability, accountability, better debugging, regulatory compliance.	Better decision-making, contextual understanding, faster response, reduced risk of false alarms.
Challenges	Complex explanations may be difficult to generate; trade-off between accuracy and interpretability.	Human errors, skill gap, high workload, dependence on human availability.
Goal	Explain AI decisions to make them transparent and reliable.	Create secure systems by combining AI efficiency with human expertise.

VI.ADVERSARIAL MACHINE LEARNING: THE NEW ATTACK SURFACE

Adversarial Machine Learning has become an important area of concern in cybersecurity because attackers are increasingly targeting AI systems themselves. Although AI and machine learning improve threat detection and automated defense, they also introduce new vulnerabilities that can beexploited by adversaries. Attackers use techniques such as evasion attacks, data poisoning, model extraction, and privacy attacks to manipulate AI models, cause incorrect predictions, or steal sensitive information. Large Language Models (LLMs) and generative AI systems face additional risks, including prompt injection, jailbreaking, and hidden backdoor attacks that can bypass safety controls and generate harmful outputs. To counter these threats, researchers and security professionals are developing advanced defense mechanisms such as adversarial training, defensive distillation, input preprocessing, and certified robustness techniques. Emerging adaptive defense systems that use reinforcement learning and game-theoretic approaches are also

being explored to help AI models dynamically respond to evolving attacks. Overall, adversarial machine learning highlights the need for secure, resilient, and continuously evolving AI-powered cybersecurity solutions.[4]

VII. CHALLENGES AND FUTURE RESEARCH PARADIGMS

Despite significant advancements in Artificial Intelligence (AI) and Machine Learning (ML) for cybersecurity, several challenges continue to limit their effective deployment in real-world environments. One of the major issues is the lack of high-quality and up-to-date cybersecurity datasets, as many existing datasets are outdated, incomplete, or do not accurately represent modern attack scenarios. Deep learning models also face interpretability challenges because their decision-making processes are often difficult to understand, creating trust and compliance concerns for security analysts and organizations. Additionally, AI-based security systems remain vulnerable to adversarial attacks, where attackers manipulate data or inputs to mislead models and reduce their effectiveness. Practical deployment is further complicated by the high computational cost, latency, energy consumption, and shortage of professionals with expertise in both AI and cybersecurity. Looking ahead, future research is expected to focus on privacy-preserving techniques such as federatedlearning, quantum-safe security solutions, self-healing systems capable of automatically detecting and recovering from attacks, and AI-enhanced zero-trust architectures. By combining advances in AI, cybersecurity, and data science, researchers aim to develop more robust, adaptive, and resilient cyber defense systems capable of addressing the evolving threat landscape.[5]



Fig 3 : Challenges & Future Directions

VIII. CONCLUSION

The integration of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) has transformed cybersecurity by moving from traditional rule-based methods to intelligent and proactive defense systems. These technologies can analyze large amounts of data in real time, helping organizations detect advanced cyber threats such as zero-day attacks and malware more quickly and accurately. However, challenges such as adversarial attacks and the lack of transparency in deep learning models remain important concerns. Therefore, future cybersecurity solutions should combine AI capabilities with human expertise while focusing on explainable, secure, and adaptive systems. Continued research in areas such as privacy protection, robust AI models, and self-healing security systems will help build stronger and more resilient cybersecurity frameworks for the future.[2]

REFERENCES

1. Dasgupta, D., et al. (2020). *Basics of machine learning and deep learning in cybersecurity. The Journal of Defense Modeling and Simulation.*
2. Ofusori, L., Bokaba, T., & Mhlongo, S. (2024). *Artificial intelligence in cybersecurity: A comprehensive review and future direction. Applied Artificial Intelligence, 38(1).*
3. Alqahtani, H., & Kumar, G. (2026). *Large language models for cybersecurity intelligence: A systematic review. Computers, Materials & Continua, 82(1), 1849–1880.*
4. 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.*
5. Zhang, Z., et al. (2021). *Artificial intelligence in cyber security: Research advances, challenges, and opportunities. Artificial Intelligence Review, 54(2), 1023–1053.*
6. Salem, A. H., et al. (2024). *Advancing cybersecurity: A comprehensive review of AI-driven detection techniques. Journal of Big Data, 11, Article 155.*
7. Kincl, J., Adam, M. T. P., & Pavleska, T. (2026). *Human-AI integration in cybersecurity: An industry-aligned perspective on incident management. International Journal of Information Security.*
8. Charmet, F., et al. (2022). *Explainable artificial intelligence for cybersecurity: A literature survey. Annals of Telecommunications, 77(11–12), 789–812.*
9. Ahsan, M., et al. (2022). *Cybersecurity threats and their mitigation approaches using machine learning—A review. Journal of Cybersecurity and Privacy, 2(3), 527–555.*
10. Sahu, R., et al. (2025). *AI-driven anomaly detection in cyber-physical power systems. Scientific Reports, 15, Article 23078.*

