MULTIDISCIPLINARY RESEARCH INSTITUTE INDIA

Multidisciplinary Journal of Research in Engineering and Technology

ISSN: 2348-6953

Volume 13 Issue 02 (2026) | Pages: 365-369

Artificial Intelligence for Cybersecurity: Threats, Attacks and Mitigation — A Systematic Literature Review

Shweta Meshram[1], A. K. Kankatre[2], V. K. Girjapure[3], A. M. Dhadage[4], K. J. Nayakawadi[5], S. P. Dakre[6]

[1,2,3,4,5,6]Department of MCA, MES's Institute of Management and Career Courses, Pune

[1]sdm.imcc@mespune.in, [2]kankatreatharv@gmail.com, [3]vedantgirjapure40@gmail.com, [4]akashdhadage24@gmail.com, [5]kunalnayakawadi1010@gmail.com, [6]shravandakare23@gmail.com

Peer Review Information

***Type**: Article*
***Received**: 20 March 2026*
***Revised**: 03 April 2026*
***Accepted**: 21 May 2026*
***Published**: 03 June 2026*

Abstract

In terms of the current landscape, cybersecurity is a multifaceted and dynamic tapestry that requires comprehensive understanding and vigilance. While AI has the potential to revolutionize cybersecurity, it is crucial to recognize the potential risks associated with its application. The integration of AI in cybersecurity can provide pivotal benefits, such as enhanced threat detection, proactive security measures, and improved incident response. However, AI can also be leveraged by cybercriminals to develop more sophisticated and intelligent attacks. As such, it is essential to adopt a strategic and balanced approach, striking a balance between exploiting its potential benefits and mitigating the associated risks.

Keywords: Cybersecurity; Cyber-attacks; DDoS; Man-in-the-Middle; Intrusion Detection; Artificial Intelligence.

How to Cite This Article

Meshram, S., Kankatre, A. K., Girjapure, V. K., Dhadage, A. M., Nayakawadi, K. J., & Dakre, S. P. (2026). Artificial intelligence for cybersecurity: Threats, attacks and mitigation—A systematic literature review. *Multidisciplinary Journal of Research in Engineering and Technology*, 13(2), 365–369.

Introduction

While the rapid adoption of digital technologies in all facets of human life has brought about many benefits, it has also opened the door for more sophisticated cybersecurity risks. The frequency, sophistication, and destructiveness of cyberattacks like ransomware, phishing, data breaches, and denial-of-service (DoS) attacks have increased dramatically. Traditional cybersecurity defenses — which are frequently static and rule-based — have proven inadequate against dynamic and intelligent threats.

AI is becoming a disruptive force in cybersecurity because of its capacity to learn from data, identify patterns, and adapt to novel circumstances. Businesses can improve their ability to detect, respond to, and prevent threats by integrating AI technologies into their cyber defense systems. This systematic literature review (SLR) examines the use of AI to improve cybersecurity, highlighting key AI methods, applications, challenges, and potential research directions.

Research Questions (RQS)

To guide this review, the following research questions were formulated:

- RQ1: What are the key AI techniques utilized in the cybersecurity domain?
- RQ2: How is AI applied to enhance cybersecurity measures in various areas?
- RQ3: What challenges and limitations are associated with integrating AI?
- RQ4: What future research directions can enhance the effectiveness of AI in cybersecurity?

Research Methodology (SLR)

Data Sources

A variety of academic databases were searched including IEEE Xplore, SpringerLink, ACM Digital Library, Elsevier ScienceDirect, and Google Scholar.

Search Strategy

Search terms included "AI in cybersecurity," "deep learning network security," "machine learning cyber defense," "AI intrusion detection," and "cyber threat mitigation with AI." Boolean operators were used to refine and combine terms.

Inclusion / Exclusion Criteria

- Inclusion: Peer-reviewed articles (2016–2022), English language, AI-focused cybersecurity applications.
- Exclusion: Non-peer-reviewed, non-English, or purely traditional (non-AI) methods.

Data Extraction and Synthesis

A standardized form extracted: study goals, AI methods, application domain, results, and limitations. Results were compared using a narrative synthesis approach.

AI Techniques in Cybersecurity (RQ1)

AI encompasses a diverse variety of computational methods that have been successfully applied across multiple cybersecurity domains, each offering distinct advantages.

- Machine Learning (ML): Without explicit programming, ML allows systems to learn from data and improve over time. Supervised learning algorithms such as SVM and decision trees are frequently used in spam and phishing detection, while unsupervised techniques like clustering identify anomalies in network traffic patterns.
- Deep Learning (DL): As a subset of machine learning, DL models intricate patterns using multilayer neural networks. It has demonstrated remarkable success in network traffic analysis, attacker behavior prediction, and automated malware classification at scale.
- Natural Language Processing (NLP): NLP processes and interprets human-language data to identify malicious content, phony URLs, and phishing emails. It enables automated parsing of threat reports, vulnerability disclosures, and dark-web chatter to provide early warning of emerging attack campaigns.
- Expert Systems: Expert systems replicate human decision-making through the application of a set of rules to a knowledge base. They are employed in IDS to appraise threats based on predefined rules and expert knowledge, offering transparent and explainable decisions.

- Reinforcement Learning (RL): RL trains agents to discover effective defensive actions through reward-based trial-and-error interaction with a simulated environment. It is increasingly used to automate threat response, generate adversarial test cases, and adapt defenses dynamically against novel attack strategies.

Applications Of AI In Cybersecurity (RQ2)

The integration of AI is increasingly prevalent within the cybersecurity domain, serving to augment threat identification, reaction, and mitigation across multiple layers.

- Intrusion Detection Systems (IDS): AI significantly improves IDS by enhancing the detection of anomalous activity in network traffic. High-accuracy identification of possible intrusions is achieved through ML algorithms including SVM, k-means clustering, and deep neural networks.

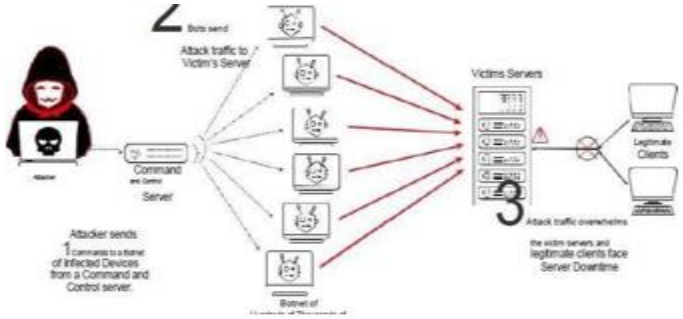

Fig. 1. *Distributed Denial of Services Attack scenario.*

- Malware Detection: AI-based systems identify malware variants by learning behavioral signatures from previously observed samples, examining source code, binary files, and runtime patterns.
- Phishing and Spam Detection: NLP and ML models identify phishing attempts by analyzing suspicious links, sender metadata, and deceptive linguistic patterns in emails and malicious websites.
- Behavioral Analysis: Continuous user behavior monitoring allows AI to flag irregularities that indicate compromised accounts or insider threats; behavioral biometrics enable ongoing passive authentication without interrupting the user's workflow.
- Risk Assessment and Management: AI models forecast weaknesses by examining threat intelligence feeds, system configurations, and historical data, enabling proactive risk mitigation.
- Fraud Detection: In banking and e-commerce, AI identifies anomalous patterns in transaction histories and user behavior, helping detect fraudulent transactions in real time.

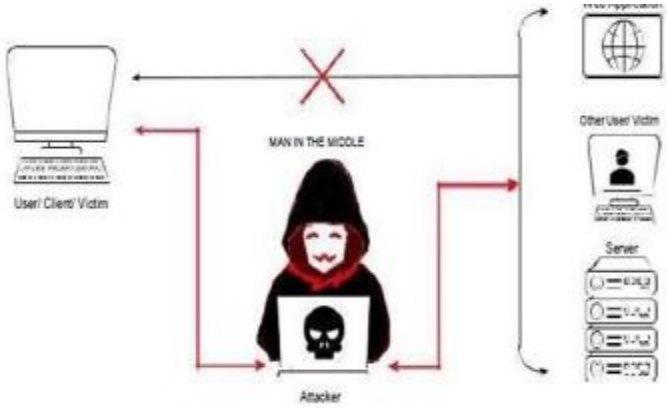

Fig. 2. *A typical illustration of SQL injection attack.*

Challenges And Limitations (RQ3)

- Data Requirements: AI models require large amounts of labeled data for accurate training, but collecting such data is difficult due to privacy concerns and limited availability.
- Adversarial Attacks: Adversarial attacks manipulate input data to deceive AI models, causing them to produce incorrect or misleading results that could compromise security systems.

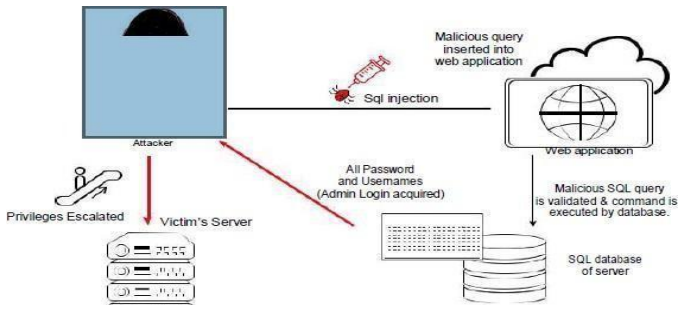

Fig. 3. *A simple depiction of Man in the Middle Attacks.*

- Model Interpretability: Many AI models, especially deep learning systems, function as black boxes, making their decision-making process difficult to understand, audit, and explain to stakeholders.
- Infrastructure and Computational Cost: Training and deploying AI models require high computational power and advanced infrastructure, making them expensive and resource-intensive for many organizations.
- Fairness and Bias: AI systems can produce biased or unfair outcomes if training data contains inherent biases, negatively affecting accuracy and reliability of security decisions.

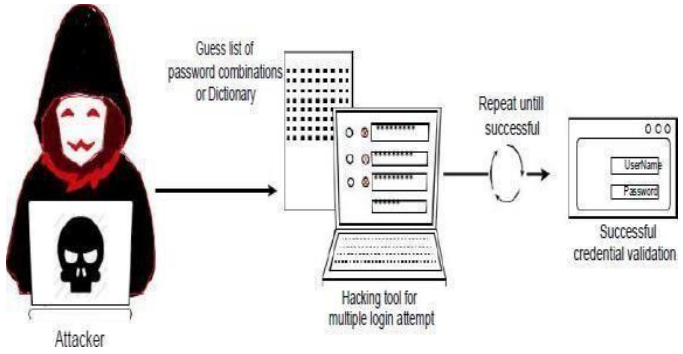

Fig. 4. *Depiction of generalised principle of a Password Attack.*

Table 1. Comparison Of AI Techniques

Technique	**Pros**	**Cons**
Machine Learning	Interpretable; fast training	Requires labeled data
Deep Learning	High accuracy; handles complex data	High resource consumption
Natural Language Processing (NLP)	Effective for text analysis	May misread context
Expert Systems	Rule-driven; explainable	Inflexible; hard to scale
Reinforcement Learning	Adaptive; self-improving	Complex; time-consuming

Future Directions (RQ4)

- Explainable AI (XAI): Explainable AI makes AI systems more transparent by clearly describing how decisions are made, improving trust and helping experts verify and validate AI outputs in critical cybersecurity applications.
- Federated Learning: Federated learning trains AI models on decentralized data without transferring it to a central server, enhancing data privacy, reducing breach risk, and ensuring compliance with data protection regulations.
- Blockchain Integration: Blockchain integration with AI provides a tamperproof and transparent system, ensuring data integrity in AI training and decision-making, and preventing unauthorized access or manipulation of security-critical data.
- Real-time Threat Intelligence: Real-time AI systems analyze data instantly as it is generated. Low-latency processing ensures quick alerts and proactive defense mechanisms, essential for preventing large-scale cyber attacks and minimizing damage.
- IoT Security: With rapid IoT growth, lightweight AI models detect threats on resource-constrained devices, monitoring activities and identifying anomalies in real time to protect interconnected systems and critical infrastructure from cyberattacks.

Conclusion

Artificial intelligence is transforming cybersecurity by making systems smarter and more capable of handling threats automatically. It enables faster attack detection and more efficient response compared to traditional methods.

AI-powered IDS, malware detectors, phishing filters, and behavioral analytics are now indispensable tools in modern security operations centers.

Significant challenges remain: the need for large labeled datasets, opacity of deep learning models, and adversarial attacks targeting AI systems themselves represent serious obstacles that must be addressed before AI can be fully trusted in realworld deployments. Issues of fairness, bias, and computational cost also require attention.

Promising future directions include Explainable AI (XAI) for transparency, federated learning for privacy, blockchain for data integrity, and lightweight models for IoT environments. With continuous research, AI has the potential to make cybersecurity stronger and more reliable worldwide.

Acknowledgements

The authors express their sincere gratitude to Dr. Shweta Meshram for her expert guidance and constructive feedback throughout this research. Special thanks are extended to the faculty and staff of MES's Institute of Management and Career Courses for providing the academic resources and infrastructure that supported this systematic literature review. The authors also acknowledge the contributions of the broader cybersecurity research community whose published work formed the foundation of this review.

Author Contributions

All five authors contributed equally to the formulation of research questions, literature search, data extraction, and synthesis. Atharv Kankatre and Vedant Girjapure led the review of

AI technique literature. Akash Dhadage and Kunal Nayakawadi conducted the application and challenges analysis. Shravan Dakre coordinated the future directions section and final manuscript preparation. All authors reviewed and approved the final manuscript.

References

1. W. Al-Yaseen, Z. Othman, and M. Z. Ahmad Nazri,"Multilevel hybrid SVM and extreme learning machine based on modified k-means for IDS," *Expert Systems with*
2. *Applications*, vol. 67, 2017.
3. I. Baptista, S. Shiaeles, and N. Kolokotronis, "A novelmalware detection system based on ML and binary visualization," in *Proc. IEEE ICCW*, 2019.
4. M. Chowdhury, A. Rahman, and M. R. Islam, "Malwareanalysis and detection using data mining and ML classification," Springer, pp. 266-274, 2018.
5. S. Coull and C. Gardner, "Activation analysis of abytebased deep neural network for malware classification," in *Proc. IEEE SPW*, pp. 21-27, 2019.
6. L. Demetrio et al., "Explaining vulnerabilities of deeplearning to adversarial malware binaries," 2019.
7. F. Feng et al., "The application of a novel neural networkin the detection of phishing websites," *J. Ambient Intelligence and Humanized Computing*, 2018.
8. W. Feng et al., "A support vector machine based naive Bayes algorithm for spam filtering," in *Proc. IEEE 35th Int.*
9. *Performance Computing and Communications Conf. (IPCCC)*, 2016.
10. H. Hashemi et al., "Graph embedding as a new approachfor unknown malware detection," *Journal of Computer Virology and Hacking Techniques*, vol. 13, pp. 153-166, 2017.
11. R. Mahajan and I. Siddavatam, "Phishing websitedetection using machine learning algorithms," *Int. Journal of Computer Applications (IJCA)*, vol. 181, no. 24, 2018.
12. Y. Ye, T. Li, D. Adjeroh, and S. S. Iyengar, "DeepAM: aheterogeneous deep learning framework for intelligent malware detection," *Knowledge and Information*
13. *Systems*, vol. 54, no. 2, pp. 265-285, 2018.
14. Cybersecurity Ventures, "Annual Cybercrime Report,"Cybersecurity Ventures, Sausalito, CA, USA, 2022. [Online].
15. Available: https://cybersecurityventures.com/
16. World Economic Forum, "Global Cybersecurity Outlook2022," WEF Insight Report, Geneva, Switzerland, Jan.
17. 2022. [Online]. Available: https://www3.weforum.org/docs/