Role of Artificial Intelligence in Cybersecurity: Strengths, Challenges, and Future Directions

Himashaili Donavalli
Dept. of Computer Science
Middle Tennessee State University
Murfreesboro, USA
hd3q@mtmail.mtsu.edu

Chase Parish
Dept. of Computer Science
Middle Tennessee State University
Murfreesboro, USA
cp6q@mtmail.mtsu.edu

***Abstract*—As cyberattacks become more advanced and harder to detect, traditional security methods like firewalls and antivirus software are no longer enough to protect sensitive systems. This review paper explores how Artificial Intelligence (AI) is transforming cybersecurity by enabling faster, smarter, and more adaptive threat detection. It focuses on key AI techniques such as machine learning, deep learning, natural language processing, and anomaly detection, highlighting how they help detect malware, phishing, and insider threats more effectively. The paper also compares AI-based approaches with traditional ones, discusses the strengths and limitations of each technique, and reviews recent studies published between 2015 and 2025. While AI offers many advantages, it also brings challenges like adversarial attacks, high data needs, and lack of transparency. The paper concludes by identifying future directions for building safer and more reliable AI systems to strengthen cybersecurity.**

I. Introduction

In recent years, cyberattacks have grown not only in number but in severity, targeting everything from small businesses to national infrastructure. High-profile incidents like the 2021 Colonial Pipeline ransomware attack, which disrupted fuel supply across the eastern United States, or the SolarWinds breach that affected several U.S. government agencies, show how devastating the consequences can be. These attacks often result in massive financial losses, reputational damage, and compromise of sensitive data. Despite efforts to strengthen digital defenses, attackers are constantly evolving—outpacing traditional cybersecurity methods that rely on known threat signatures and static rules.

Cybersecurity has become an essential part of modern software development, especially as software systems grow more complex and interconnected. Traditional security practices often detect vulnerabilities too late in the process, allowing risks to go unnoticed until systems are deployed. Andrade et al. [1] address this issue by introducing the use of large language models (LLMs) within DevSecOps pipelines to provide real time alerts on insecure coding practices. This proactive approach helps developers identify and resolve security flaws early, strengthening the software before it even reaches production. By integrating AI into development workflows, this method reduces the chances of future attacks and builds more secure software from the ground up.

As cyber threats grow more frequent and advanced, traditional security tools like firewalls and antivirus software struggle to keep up. This has led to increased adoption of Artificial Intelligence (AI) in cybersecurity. Siam et al. [2] provide a comprehensive overview of how AI through techniques such as machine learning (ML), deep learning (DL), natural language processing (NLP), and anomaly detection is transforming the cybersecurity landscape. Unlike conventional systems that rely on known threat signatures, AI based tools can analyze large volumes of data, detect unusual patterns, and adapt to new threats in real time. For example, tools like Darktrace use unsupervised learning to identify network anomalies, while IBM Watson leverages NLP to process and understand threat intelligence. These AI systems have also been used to detect phishing, insider threats, and malware with improved speed and accuracy.

However, the use of AI in cybersecurity also introduces new challenges. AI models require vast amounts of high quality data and are often considered "black boxes," making their decision-making difficult to interpret. They are also vulnerable to adversarial attacks, where malicious actors craft deceptive inputs to fool the system. Siam et al. further point out that while individual AI techniques have their own strengths, no single approach is sufficient on its own. Hybrid models that combine multiple AI strategies may offer more robust protection.

Looking deeper into the potential risks of relying on AI, Yampolskiy and Spellchecker [3] highlight a series of alarming real world failures where AI systems made dangerous or incorrect decisions. In the context of cybersecurity, such errors could result in significant damage. The authors emphasize that AI safety must go beyond simply reducing attacks; it must aim for zero failures especially as we develop increasingly powerful systems like Artificial General Intelligence (AGI). They also raise concerns about the emergence of intentionally harmful AI, which could pose even greater risks than accidental failures. The paper stresses the importance of applying cybersecurity principles not just to protect systems from attackers, but also to safeguard against the AI systems themselves.

This paper aims to explore how AI is reshaping cybersecurity starting from early development integration to real-time threat detection and the safety concerns of advanced AI systems. By reviewing current techniques, their limitations, and future directions, this study provides a balanced understanding of AI's evolving role in securing the digital world.

II. Research Design

A. Study Design

This review paper focuses on how Artificial Intelligence (AI) is being used to enhance cybersecurity, especially in areas where traditional methods like firewalls and antivirus tools fall short. The study compares both approaches and highlights how AI techniques—such as machine learning (ML), deep learning (DL), natural language processing (NLP), and anomaly detection—help in detecting cyber threats more accurately and efficiently. It also examines the limitations and risks of AI in cybersecurity and suggests future directions for improvement.

B. Search Strategy

A structured literature search was done using trusted databases: *IEEE Xplore, Google Scholar, ResearchGate, SpringerLink, Scopus, Pubmed*. The search covered studies published between 2015 and 2025 in English. Boolean operators (AND/OR) were used to improve search accuracy. Reference lists of selected papers were also reviewed to find more relevant studies.

The keywords used include: *"AI in cybersecurity", "Machine learning for threat detection", "Deep learning in cybersecurity", "NLP in phishing detection", "Anomaly detection in cyber attacks", "AI vs traditional cybersecurity methods", "Challenges of AI in security"*.

C. Inclusion and Exclusion Criteria

Inclusion criteria: This review included peer-reviewed papers published between 2015 and 2025 that focus on AI techniques in cybersecurity. The selected studies specifically addressed the role of AI methods like ML, DL, NLP, and anomaly detection in cybersecurity applications. Preference was given to papers that compared AI with traditional security approaches and clearly discussed both the advantages and limitations of AI-driven methods.

Exclusion criteria: Studies that did not involve AI in any form of cybersecurity application were excluded. Additionally, papers that focused only on tools without explaining the underlying AI methods were not considered. Highly technical or purely mathematical papers without practical relevance were also excluded. Non-peer-reviewed content such as blogs, editorials, or opinion pieces were filtered out during the screening process.

D. PRISMA Flowchart and Selection of Studies

The study selection process followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach to ensure a clear and systematic review. A total of 29 records were identified from databases including IEEE Xplore, Scopus, and Google Scholar. After reviewing titles and abstracts, 21 records were shortlisted for screening. Among these, 8 records were excluded as they did not meet the research focus or relevance criteria.

The remaining 21 papers were assessed for eligibility through full-text review. Of those, 10 were excluded because they did not align with the required focus on AI techniques or lacked sufficient discussion of cybersecurity relevance. Finally, 11 studies were included in the review. These selected studies form the basis of the analysis and findings discussed in later sections. The PRISMA process is shown in Figure 2.

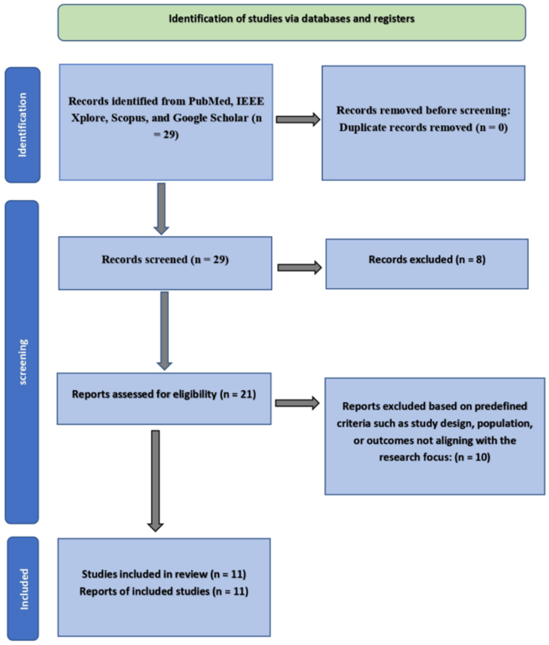

Identification of studies via databases and registers
Identification
Records identified from PubMed, IEEE Xplore, Scopus, and Google Scholar (n = 29)
Records removed before screening: Duplicate records removed (n = 0)
screening
Records screened (n = 29)
Records excluded (n = 8)
Reports assessed for eligibility (n = 21)
Reports excluded based on predefined criteria such as study design, population, or outcomes not aligning with the research focus: (n = 10)
Included
Studies included in review (n = 11)
Reports of included studies (n = 11)

Fig. 1. PRISMA flowchart showing the selection process of studies

E. Data Extraction

The data extraction process involved systematically reviewing each of the 11 selected studies to gather key information relevant to the research questions. For every paper, the author(s) and publication year were noted. The specific AI method used—such as supervised learning, deep learning, NLP, or anomaly detection—was identified and categorized.

Next, the cybersecurity challenge addressed by each AI technique was recorded. These included common threat types like phishing, malware attacks, and insider threats. For each study, strengths of the AI method were noted, such as fast detection times, the ability to analyze large datasets, or adaptability to new threats.

Alongside the benefits, each study's reported limitations were also extracted. Common challenges included vulnerability to adversarial attacks, dependency on large high-quality datasets, and the lack of explainability in some AI models. If a study included comparisons to traditional cybersecurity systems, those comparisons were recorded as well. All collected data were then grouped based on technique type and cybersecurity function, enabling a clear thematic analysis in the next section.

F. Data Synthesis

The extracted data were organized into four key themes:

AI vs. Traditional Security Methods: Traditional cybersecurity tools rely on fixed rules and known signatures, which limits their ability to detect new or hidden threats. In contrast, AI techniques can process vast datasets, detect novel attack patterns, and adapt to changing environments. Across the reviewed studies, AI consistently outperformed traditional methods in detecting modern cyber threats and produced fewer false positives.

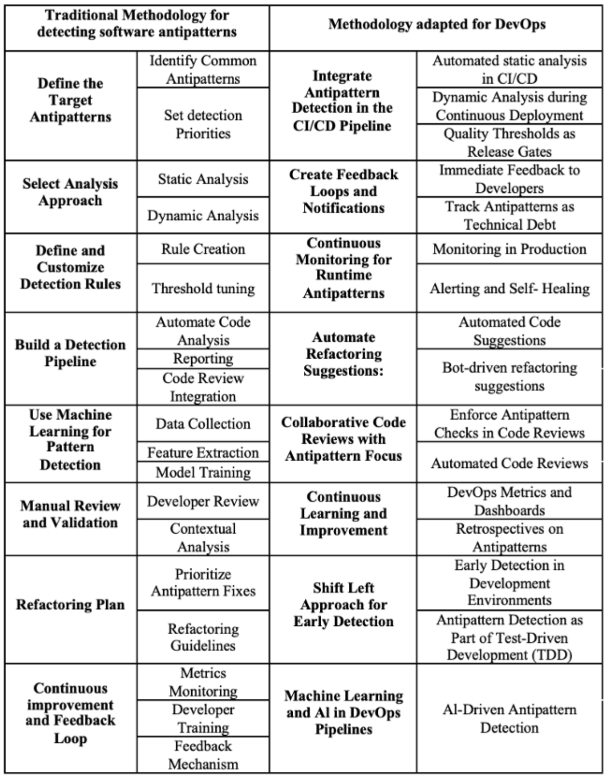
Role of AI Techniques in Cybersecurity: Each AI technique reviewed played a unique role. Machine Learning (ML) detected patterns in historical data and flagged suspicious behavior. Deep Learning (DL) handled complex and unstructured data, often used in advanced malware or anomaly detection. Natural Language Processing (NLP) identified phishing content and processed threat intelligence in textual form. Anomaly Detection spotted deviations from normal system behavior that might indicate a cyberattack.

Strengths and Challenges: Strengths of AI included real-time threat detection, adaptability to new attack types, and scalability across large datasets. Challenges involved vulnerability to adversarial inputs, reliance on high-quality labeled data, and opacity in decision-making (i.e., black-box models).

Future Directions: To address existing limitations, future efforts should focus on building hybrid models that combine multiple AI techniques for better accuracy and coverage. There is a growing need for explainable AI to make security decisions transparent and trustworthy. Researchers also emphasize the importance of designing adversarially robust AI systems and using privacy-preserving methods like federated learning. Lastly, ethical considerations and responsible AI development will play a key role in shaping secure and reliable AI-driven cybersecurity in the years ahead.

III. Results

From the reviewed literature, it is clear that traditional cybersecurity tools like Static Application Security Testing (SAST) and Dynamic Application Security Testing (DAST) are limited in their ability to detect modern threats. These tools rely on fixed rules and known attack signatures, which makes them less effective at identifying new or complex vulnerabilities. In contrast, Artificial Intelligence (AI), especially models like Large Language Models (LLMs), can understand the context of code, learn from vast datasets, and detect hidden issues much earlier in the development process. Andrade et al. [1] demonstrated that AI-based methods, when integrated into DevOps pipelines, can identify software antipatterns in real time—something that traditional scanners often miss. Their study also showed in Figure 2 a comparison between a standard software workflow and an AI-enhanced DevOps pipeline, highlighting how the AI model improves security during continuous integration and deployment phases. Additional studies by Pearce et al. and Purba et al. [4][5] further support this, showing that LLMs not only detect vulnerabilities using reasoning and analogy but also reduce false positives, which is a common issue in traditional tools. Real-world tools such as GitHub Copilot and Microsoft Security Copilot have demonstrated these capabilities. GitHub Copilot provides developers with real-time code suggestions across languages [6], while Microsoft Security Copilot, powered by GPT-4, helps teams analyze and prioritize threats collaboratively and effectively in live environments [7]. Overall, the reviewed results consistently highlight how AI techniques outperform traditional methods in terms of accuracy, speed, and adaptability when detecting modern cybersecurity threats.



Traditional Methodology for detecting software antipatterns		**Methodology adapted for DevOps**	
Define the Target Antipatterns	Identify Common Antipatterns Set detection Priorities	**Integrate Antipattern Detection in the CI/CD Pipeline**	Automated static analysis in CI/CD Dynamic Analysis during Continuous Deployment Quality Thresholds as Release Gates
Select Analysis Approach	Static Analysis Dynamic Analysis	**Create Feedback Loops and Notifications**	Immediate Feedback to Developers Track Antipatterns as Technical Debt
Define and Customize Detection Rules	Rule Creation Threshold tuning	**Continuous Monitoring for Runtime Antipatterns**	Monitoring in Production Alerting and Self- Healing
Build a Detection Pipeline	Automate Code Analysis Reporting Code Review Integration	**Automate Refactoring Suggestions:**	Automated Code Suggestions Bot-driven refactoring suggestions
Use Machine Learning for Pattern Detection	Data Collection Feature Extraction Model Training	**Collaborative Code Reviews with Antipattern Focus**	Enforce Antipattern Checks in Code Reviews Automated Code Reviews
Manual Review and Validation	Developer Review Contextual Analysis	**Continuous Learning and Improvement**	DevOps Metrics and Dashboards Retrospectives on Antipatterns
Refactoring Plan	Prioritize Antipattern Fixes Refactoring Guidelines	**Shift Left Approach for Early Detection**	Early Detection in Development Environments Antipattern Detection as Part of Test-Driven Development (TDD)
Continuous improvement and Feedback Loop	Metrics Monitoring Developer Training Feedback Mechanism	**Machine Learning and AI in DevOps Pipelines**	AI-Driven Antipattern Detection

Fig. 2. Traditional and DevOps methodology for detecting software antipatterns [1]

One of the most important roles identified in the reviewed studies is that of Machine Learning (ML) in identifying cyber threats by learning from past patterns. Unlike traditional rule-based systems, ML models can adapt over time to recognize new behaviors. Supervised learning methods like decision trees and support vector machines (SVM) are commonly used to classify network activities as normal or malicious, using labeled datasets Sheykhmousa et al [8] Unsupervised models, on the other hand, detect unknown threats by analyzing network traffic or user activity without prior labeling. These methods help detect unusual events and potential attacks in real time. In cybersecurity, semi-supervised learning also plays a role by using small amounts of labeled data with larger unlabeled data, helping to identify advanced threats more efficiently. These ML methods form the foundation for behavior analytics and threat detection systems Al Siam et al [2].

The review also emphasizes the importance of Deep Learning (DL) in detecting more complex and hidden cyber threats, especially from unstructured data like raw files or sequential

TABLE I
SUMMARY OF AI TECHNIQUES IN CYBERSECURITY

Model	Strengths	Primary Applications	Limitations
Machine Learning (ML)	Learns from past data and adapts over time to recognize new patterns. Provides real-time behavior analysis.	Used for detecting threats and analyzing user or system behavior based on supervised, unsupervised, or semi-supervised learning.	May struggle with identifying new threats without retraining. Needs regular updates to remain effective.
Deep Learning (DL)	Highly accurate when analyzing complex and unstructured data like files and behavior logs.	Effective in detecting malware and phishing by analyzing sequences or raw data using CNNs, RNNs, and autoencoders.	Requires high computing power and is difficult to interpret (black-box models).
Natural Language Processing (NLP)	Good at detecting threats in text data. Helps in fast and automated analysis of emails, documents, and messages.	Used for phishing detection, content classification, and extracting threat intelligence from open-source data (OSINT).	May face challenges with multilingual content and domain-specific terms.
Anomaly Detection Algorithms	Suitable for real-time monitoring; builds baselines and detects deviations from normal behavior.	Applied in intrusion detection and spotting insider threats using statistical or ML-based models.	Can generate false positives in dynamic or frequently changing environments.

user activity. DL models such as Convolutional Neural Networks (CNNs) are effective in identifying malware patterns even when files are disguised using obfuscation techniques Gibert et al [9]. Recurrent Neural Networks (RNNs), which process data over time, are used to track user behavior or detect suspicious activities across long sessions, such as insider threats or slow attacks. Additionally, autoencoders are helpful in learning what ”normal” behavior looks like and flagging anything unusual as a threat. While DL models provide very accurate detection, they also require more computing power and are harder to explain. Still, they offer a strong advantage in detecting modern, evolving threats Al Siam et al [2].

In terms of text-based threats, Natural Language Processing (NLP) plays a key role in cybersecurity, especially for identifying phishing and social engineering attacks. NLP models can read and analyze emails, chat logs, and documents to find unusual or risky content. For example, phishing detection is done by analyzing the words, tone, and sender information in messages that try to trick users into revealing sensitive information Balantrapu [10]. NLP also helps categorize content using models like TF-IDF or BERT to flag harmful links or documents automatically. Additionally, techniques like semantic analysis and entity recognition help detect private data and the intent behind a message. These tools are especially useful in analyzing large volumes of open source intelligence and email communication Al Siam et al [2].

Finally, the reviewed results show that Anomaly Detection algorithms are essential for identifying unusual system or network behavior that may indicate a cyberattack. These methods establish a baseline for what is considered “normal” and flag deviations. Statistical models like Z-score calculations are used to detect outliers, while algorithms such as k-nearest neighbors (k-NN) measure the distance between activity patterns to identify irregularities Domanski [11]. Machine learning–based models like Isolation Forests and One Class SVMs are particularly good at identifying threats that do not match known patterns. Although anomaly detection is useful for early intrusion alerts, it may produce false positives in rapidly changing environments. Still, it offers strong value in real-time threat monitoring Al Siam et al [2].

A. *Strengths of AI in Cybersecurity*

The reviewed studies highlight that one of the major strengths of Artificial Intelligence (AI) in cybersecurity is its ability to detect threats in real time and scale across large systems. As cyberattacks become more complex and faster, AI tools can monitor vast amounts of data and respond quickly to unusual activities. This makes AI especially effective in identifying attacks that happen too quickly or subtly for human operators to catch. According to Yampolskiy and Spellchecker [3], AI’s rapid processing and automated decision-making capabilities provide a strong advantage in environments where traditional systems may lag or overlook subtle anomalies. AI’s learning ability also allows it to adapt to new attack types over time something static, rule-based systems struggle with. When deployed across distributed networks, AI models can maintain high performance even at large scale, which makes them useful in both enterprise and national cybersecurity settings.

B. *Challenges of AI Systems in Security*

Despite these strengths, several serious challenges remain with using AI in cybersecurity. One key issue is that AI systems are often considered ”black boxes,” meaning their decision-making process is hard to understand or explain. This makes it difficult for security teams to trust or validate AI generated alerts. The paper points out that such opacity can be dangerous especially in high-risk scenarios where wrong decisions might cause more harm than help Yampolskiy and Spellchecker[3]. Another major challenge is AI’s vulnerability to adversarial attacks. If attackers deliberately feed misleading inputs, AI systems might fail or behave unexpectedly. Moreover, many AI models depend heavily on large, high-quality labeled datasets. If the data used for training is biased, incomplete, or outdated, the AI’s decisions will also be flawed. As noted by the authors, even well designed systems have failed in the past due to poor training assumptions or unexpected real world inputs, such as facial recognition software misclassifying people or trading bots causing financial crashes [12].

C. Future Directions for AI in Cybersecurity

To overcome these limitations, researchers propose several future directions. First, combining multiple AI models such as integrating machine learning with rule-based logic or using ensemble methods—can improve accuracy and reduce risk. Yampolskiy emphasizes the need for "AI Safety Engineering", a structured field focused on preventing failures before they occur. One promising direction is explainable AI (XAI), which aims to make AI decisions more understandable and trustworthy for human analysts. Another area is federated learning, which allows AI models to learn from distributed data without compromising privacy important in sensitive environments like healthcare or defense. The paper also stresses the importance of making AI systems robust against adversarial attacks, by designing models that can detect and resist manipulation attempts. Finally, ethical design and responsible AI development including transparency, accountability, and safety testing are identified as crucial for ensuring AI does not unintentionally cause harm as it becomes more integrated into our digital infrastructure.

IV. Discussion and Observations

The results of this review reveal that Artificial Intelligence (AI) is playing a transformative role in cybersecurity. AI techniques such as machine learning (ML), deep learning (DL), natural language processing (NLP), and anomaly detection have proven to be far more effective than traditional rule-based tools when it comes to detecting complex and modern threats. Unlike conventional systems that rely on fixed signatures, AI can process large datasets, learn from past behaviors, and detect new or subtle threats in real time. One important observation is the increasing use of Large Language Models (LLMs) in development pipelines (Andrade et al., 2024), which allow insecure code patterns to be identified during software development rather than after deployment. This proactive security approach is a noticeable shift from traditional reactive methods.

An unexpected but important finding is the wide range of tasks AI can handle across different threat types. For example, while ML is used for behavior monitoring, NLP is effectively spotting phishing and textual social engineering, and DL is helping identify hidden threats in unstructured data like network traffic or logs (Siam et al., 2025). Interestingly, many AI systems were found to outperform human analysts in speed and volume of analysis, but they lacked transparency, which raised concerns about decision validation. In some cases, reviewed studies also noted overlapping functionalities—for instance, both DL and anomaly detection were used to uncover abnormal behavior, suggesting possible integration potential. Additionally, the literature shows a growing trend toward hybrid models, where multiple AI techniques are combined to improve accuracy, reduce false positives, and cover more threat scenarios (Yampolskiy and Spellchecker, 2016).

V. Limitations

Despite AI's potential in cybersecurity, the review highlights some notable limitations. First, many of the AI models discussed require large, high-quality labeled datasets to perform well. In real-world environments, such data may not always be available or balanced, making model training harder. Second, there is a common concern about the "black-box" nature of AI systems—where it is not clear how or why certain decisions are made. This lack of interpretability makes it difficult for security analysts to trust or act on AI alerts, especially in high-risk contexts (Yampolskiy and Spellchecker, 2016). Another issue is vulnerability to adversarial inputs: attackers can manipulate data in ways that confuse or mislead the model, causing it to fail silently. Moreover, the studies reviewed were limited in terms of long-term empirical deployment. Most findings were based on experimental data or short-term evaluations, which leaves open questions about long-term reliability and scalability in real-life security operations. Time constraints and tool-specific limitations also affected some research results.

VI. Conclusion

This review paper explored how Artificial Intelligence is reshaping cybersecurity by analyzing its techniques, benefits, and limitations. The results showed that AI methods offer significant advantages over traditional tools, especially in real-time detection, adaptability to new threats, and large-scale data processing. The use of LLMs and hybrid models in DevOps pipelines marks a key advancement in early threat detection. However, challenges such as data dependency, lack of explainability, and adversarial vulnerabilities remain critical. Overall, the answer to the research question—"Can AI enhance cybersecurity beyond traditional methods?"—is yes, but with important caveats. For future research, there is a need to focus on building explainable and ethically designed AI systems, testing models against adversarial attacks, and exploring privacy-preserving techniques like federated learning. The development of AI safety engineering as a structured discipline will also be vital in ensuring that as AI tools become more powerful, they remain secure, transparent, and reliable.

References

[1] Andrade, Roberto, Jenny Torres, Pamela Flores, Erick Cabezas, and Jorge Segovia. "Convergence of AI for Secure Software Development." In 2024 8th Cyber Security in Networking Conference (CSNet), 138–42. Paris, France: IEEE, 2024. https://doi.org/10.1109/CSNet64211.2024.10851473.

[2] Siam, Abdullah Al, Md Maruf Hassan, and Touhid Bhuiyan. "Artificial Intelligence for Cybersecurity: A State of the Art." In 2025 IEEE 4th International Conference on AI in Cybersecurity (ICAIC), 1–7. Houston, TX, USA: IEEE, 2025. https://doi.org/10.1109/ICAIC63015.2025.10848980.

[3] Yampolskiy, Roman V., and M. S. Spellchecker. "Artificial Intelligence Safety and Cybersecurity: A Timeline of AI Failures." arXiv, 2016. https://doi.org/10.48550/ARXIV.1610.07997.

[4] Pearce, Hammond, Benjamin Tan, Baleegh Ahmad, Ramesh Karri, and Brendan Dolan-Gavitt. "Examining Zero-Shot Vulnerability Repair with Large Language Models." In 2023 IEEE Symposium on Security and Privacy (SP), 2339–56. San Francisco, CA, USA: IEEE, 2023. https://doi.org/10.1109/SP46215.2023.10179324.

[5] Purba, Moumita Das, Arpita Ghosh, Benjamin J. Radford, and Bill Chu. “Software Vulnerability Detection Using Large Language Models.” In 2023 IEEE 34th International Symposium on Software Reliability Engineering Workshops (ISSREW), 112–19. Florence, Italy: IEEE, 2023. https://doi.org/10.1109/ISSREW60843.2023.00058.

[6] Koyanagi, Kei, Dong Wang, Kotaro Noguchi, Masanari Kondo, Alexander Serebrenik, Yasutaka Kamei, and Naoyasu Ubayashi. “Exploring the Effect of Multiple Natural Languages on Code Suggestion Using GitHub Copilot.” In Proceedings of the 21st International Conference on Mining Software Repositories, 481–86. Lisbon Portugal: ACM, 2024. https://doi.org/10.1145/3643991.3644917.

[7] Sai, Siva, Utkarsh Yashvardhan, Vinay Chamola, and Biplab Sikdar. “Generative AI for Cyber Security: Analyzing the Potential of ChatGPT, DALL-E, and Other Models for Enhancing the Security Space.” IEEE Access 12 (2024): 53497–516. https://doi.org/10.1109/ACCESS.2024.3385107.

[8] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, “Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 6308–6325, 2020.

[9] D. Gibert, C. Mateu, J. Planes, and R. Vicens, “Using convolutional neural networks for classification of malware represented as images,” Journal of Computer Virology and Hacking Techniques, vol. 15, pp. 15–28, 2019.

[10] S. S. Balantrapu, “Evaluating the effectiveness of machine learning in phishing detection,” International Scientific Journal for Research, vol. 5, no. 5, 2023.

[11] P. D. Domanski, “Study on statistical outlier detection and labelling,” International Journal of Automation and Computing, vol. 17, no. 6, pp. 788–811, 2020

[12] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, et al., ”Intriguing properties of neural networks,” arXiv preprint arXiv:1312.6199, 2013.

[13] Zhekov, Alexander. “Challenges to Ensuring Information Security and the Role of Artificial Intelligence.” In 2023 VI International Conference on High Technology for Sustainable Development (HiTech), 1–4, 2023. https://doi.org/10.1109/HiTech60680.2023.10759133.

[14] Mahmoud, Mohammed. “The Risks and Vulnerabilities of Artificial Intelligence Usage in Information Security.” In 2023 International Conference on Computational Science and Computational Intelligence (CSCI), 266–69, 2023. https://doi.org/10.1109/CSCI62032.2023.00047.

[15] Patel, Advait, Pravin Pandey, Hariharan Ragothaman, Ramasankar Molleti, and Ajay Tanikonda. “Securing Cloud AI Workloads: Protecting Generative AI Models from Adversarial Attacks.” In 2025 IEEE 4th International Conference on AI in Cybersecurity (ICAIC), 1–7, 2025. https://doi.org/10.1109/ICAIC63015.2025.10848877.