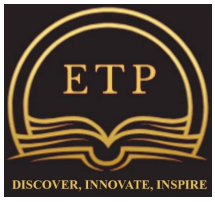
ISSN: 2710-2599



International Journal of Data Science and Big Data Analytics
Publisher's Home Page: https://www.svedbergopen.com/

Review Article Open Access

A Systematic Review of AI-Driven Approaches for Mitigating Cybersecurity

Sultanul Arifeen Hamim[1] and Akinul Islam Jony[2*]

[1]American International University, Bangladesh (AIUB), Dhaka, Bangladesh. E-mail: arifeenhamim@gmail.com
[2]Universitat Oberta de Catalunya (UOC), Barcelona, Spain. E-mail: ajony@uoc.edu

Article Info

Volume 6, Issue 1, May 2026
Received : 07 January 2026
Accepted : 28 April 2026
Published : 25 May 2026
doi: 10.67191/IJDSBDA.6.1.2026.55-67

Abstract

In the ever-evolving landscape of cybersecurity threats, Artificial Intelligence (AI) has surfaced as a pivotal defence mechanism against the rising tide of cyber assaults. This systematic review delves into the nexus of AI and cybersecurity, assessing the varied applications of AI-driven methodologies, their effectiveness, and the persistent challenges within this sphere. With the advent of AI-enabled cyber weaponry, traditional security measures are proving inadequate, demanding a shift towards more innovative, AI-based solutions. This paper evaluates the trajectory of AI development in cybersecurity, from its initial recognition to the current push for continuous innovation aimed at outpacing the sophisticated, AI-powered threats. We highlight the transformative role of AI in cybersecurity, which includes the transition to autonomous decision-making systems and the emergence of offensive AI employed by cybercriminals. Despite the potential of AI to significantly strengthen cyber defences, this review identifies a critical need for ongoing research and development to navigate the complexities of AI-driven cybersecurity threats.

Keywords: *Artificial intelligence, Cybersecurity, AI-driven security, Adaptive defense, Anomaly detection*

1. Introduction

In a digitally interconnected world, cybersecurity has become a critical issue. The rapid expansion of digital data, widespread use of smart devices, and ubiquitous internet access have brought significant advancements in convenience and efficiency. However, these developments have also increased cyber threats, casting a pervasive shadow over our online interactions. In response, the integration of AI has emerged as a powerful and promising solution to combat this continuously evolving threat landscape. This systematic review explores the intersection of AI and cybersecurity, examining how AI-driven methods are used to mitigate, detect, and

* *Corresponding author: Akinul Islam Jony, Universitat Oberta de Catalunya (UOC), Barcelona, Spain. E-mail: ajony@uoc.edu*

defend against cyber threats. Today's Cyberattacks are pervasive and often considered among the most significant tactical risks (Dixon and Eagan, 2019). These cybercrimes can lead to catastrophic financial losses, impacting individuals and organizations.

The evolving cleverness of cyberattack approaches presents an existential threat to businesses, services, and organizational infrastructures (Jony and Hamim, 2023). These attacks can disrupt business operations, erase critical data, and damage reputations. The current wave of attacks, often leveraging, surpasses human capabilities in both speed and complexity. Cybercriminals can attack at a larger scale and rapidly, and at the same time bypassing traditional detection or prevention techniques through 'offensive AI' (Guembe *et al.*, 2022). This heralds the emergence of a new, more subtle and secretive generation of cybercriminals who will significantly influence the future of cybersecurity. These emerging threats will be more intelligent, capable of making decision and taking independent action. AI's adaptability and learning capabilities are ushering in a area of sophisticated, personalized, and human-like attacks (Thanh and Zelinka, 2019). Contemporary cybercriminals utilize advanced AI techniques, which provide substantial risks to governments, companies, corporations, and individuals as well. The efficacy of conventional cybersecurity technologies is progressively diminishing in the face of these sophisticated cyber weapons (Thanh and Zelinka, 2019). AI-driven cyberattacks refer to the utilization of training robots in the art of social engineering, enabling them to execute such tactics with a degree of efficiency that surpasses that of humans. The aforementioned attacks possess the ability to adjust to the surroundings in which they infiltrate, acquiring knowledge from contextual information in order to imitate reliable aspects of the digital realm or take advantage of identified weaknesses (Guembe *et al.*, 2022). The emergence of AI-powered assaults represents more than a speculative future risk; the necessary technologies and expertise for executing such attacks are presently accessible (Dixon and Eagan, 2019). The progression of AI has resulted in substantial expansion in the realms of automation and innovation. Nevertheless, it is important to acknowledge that these AI technologies can also be employed for nefarious intentions. Recently, a discernible rise in cyberattacks propelled by AI technology. Despite the increasing adoption of secure digital environments, such as cloud storage services, organizations continue to face susceptibility to cybercriminal activities (Zouave *et al.*, 2020). The overwhelming nature of AI-driven cyberattacks poses significant challenges for cybersecurity teams in terms of their complexity and scope. The task of identifying and recruiting highly skilled cybersecurity professionals to effectively mitigate these threats is becoming progressively more difficult and costly. An AI-powered hack has the potential to leverage a wide range of cyberspace and computer resources, surpassing the capabilities of humans. This can lead to faster, more unpredictable, and highly complex attacks that may pose challenges for even the most resilient cybersecurity teams (Hamadah and Aqel, 2020). As AI gains greater efficacy in the hands of evil individuals, it is imperative for cybersecurity researchers, practitioners, and governments to devise novel solutions aimed at safeguarding cyberspace against such actors. AI-driven assaults employ sophisticated obfuscation techniques and regularly modify discernible attributes, hence potentially attaining a high degree of flexibility that renders them nearly imperceptible to conventional antivirus software. The user did not provide any text to rewrite. The intended malicious utilization of AI has been exemplified by the presence of high-security vulnerabilities and covert attacks concealed within seemingly harmless carrier apps (Mohsin and Jony, 2024). Perpetrators are consistently enhancing their tactics, with a particular emphasis on employing AI methods to intensify the impact of their actions within the realm of cyberspace, all the while ensuring their activities go unnoticed (Usman *et al.*, 2019).

This systematic review seeks to provide a wide-ranging outline of the landscape of AI-driven approaches for cybersecurity, drawing insights from a rigorous analysis of existing research and practices. It endeavors to shed light on the various AI techniques that have been harnessed to bolster cyber defenses, the effectiveness of these approaches, and the challenges that still lie ahead. By synthesizing the collective knowledge and experience in this domain, we aim to empower cybersecurity professionals, policymakers, and researchers with the knowledge necessary to navigate the complex terrain of AI-powered cybersecurity. As we embark on this exploration, we invite the reader to join us in a quest to unravel the potential, the limitations, and the promises of AI-driven cybersecurity. Together, we will traverse the digital frontier in search of innovative solutions to safeguard our digital existence in an increasingly interconnected world.

Table 1 represents a structured overview of research questions along with their justifications, specifically addressing the evaluation and differentiation of AI-driven methods in cybersecurity.

Table 1: Research Question

Research Question	Rationale
How effective are AI-driven approaches in detecting and preventing cybersecurity threats	Evaluating AI-driven cybersecurity solutions is key to determining their impact and improving security measures.
What distinguishes typical targeted assaults from AI-driven cyberattacks?	Distinguish between the main logical components of typical targeted assaults and those of AI-driven attacks.
What are the key AI techniques and algorithms employed in cybersecurity, and how do they contribute to threat detection and prevention?	Investigating the specific AI techniques used in cybersecurity helps identify the state-of-the-art methods and their applicability in addressing various types of cyber threats.

2. Background of Cybersecurity

Within the domain of digital interconnection, the field of cybersecurity arises as a complex and ever-evolving discipline, distinguished by its varied range of definitions and interpretations. The phrase 'cybersecurity' is a multifaceted concept that incorporates a wide range of practices, technologies, and principles with the objective of protecting cyberspace and its associated assets. It is important to note that perceptions of cybersecurity might vary due to subjective interpretations. The document highlights the challenge posed by the lack of a globally agreed definition of cybersecurity that adequately encompasses its multidimensional nature. This shortage of consensus impedes advancements in technology and science, underscoring the necessity for a comprehensive and inclusive comprehension of the concept (Guembe *et al.*, 2022). The field of cybersecurity is inherently multidisciplinary, extending beyond technological aspects to encompass a range of disciplines. This encompasses not solely technology remedies, but also encompasses considerations of organizational, economic, social, and political dimensions. Because cybersecurity is interdisciplinary, a holistic perspective that combines concepts from multiple disciplines is required (Thanh and Zelinka, 2019; Zouave *et al.*, 2020). The concept of 'security' in the field of cybersecurity is intricate and multifaceted, frequently characterized by individual viewpoints and ethical principles. The concept of security involves a wide range of aspects, such as physical security, human security, features of information systems, and mathematical definitions pertaining to distinct types of security. The fundamental aspect of security, however, resides in the state of being devoid of any potential harm or peril (Hamadah and Aqel, 2020; Biswas *et al.*, 2024; Kaloudi and Li, 2020).

Cybersecurity encompasses various forms of cyber threats and attacks, each with its unique implications and methodologies. Misuse of resources, or unauthorized use, involves accessing network or host resources without permission. User compromise refers to hackers obtaining sensitive information like passwords to exploit the system. A more severe form, root compromise, occurs when hackers gain access to administrator-level accounts, granting them extensive system privileges. Web compromise exploits vulnerabilities, often through methods like cross-site scripting or SQL injection, to strengthen attacks. Installed malware, a significant threat, includes various forms such as viruses that self-replicate within the system, spyware that stealthily collects user data, Trojans that mislead users into granting access, and worms that rapidly spread across networks. Denial of Service (DoS) attacks aim to disrupt services which is called DDoS (Distributed Denial of Service) attack by targeting specific hosts (Host Based), entire networks (Network Based), or using distributed methods (DDoS) involving multiple compromised systems to overwhelm the target. These threats collectively represent the multifaceted challenges in cybersecurity, necessitating robust and adaptive security measures to protect digital assets and maintain system integrity (Usman *et al.*, 2019; Ferdous *et al.*, 2024).

Cyber-attacks can be categorized into various categories based upon distinct viewpoints. The cyber-attacks in this paper are categorized (as seen in Figure 1) based on the specific impacts they have on a system or its design.

3. Methodology

This review employs a systematic and comprehensive approach to investigate various aspects of cyber security

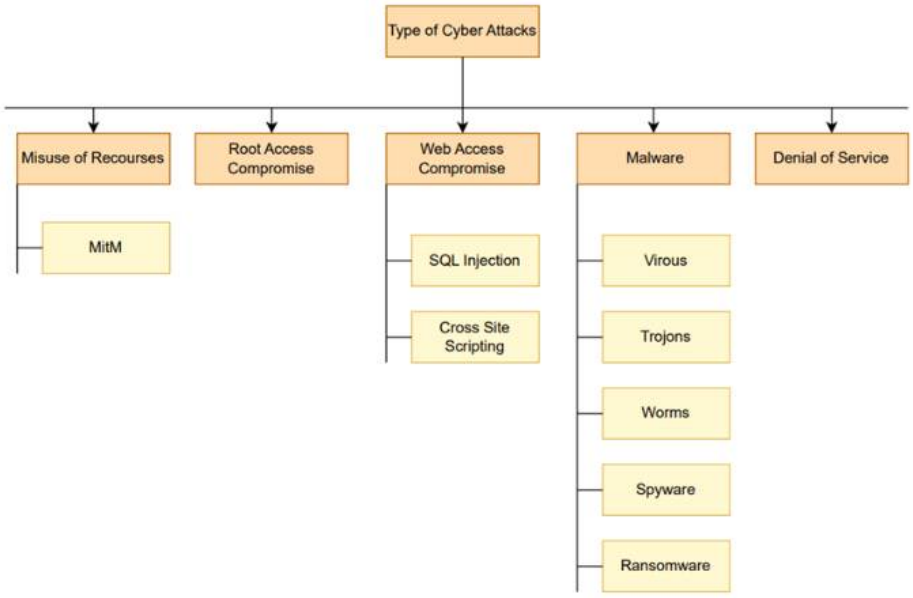


Figure 1: Types of Cyber Attacks

attacks. It is a popular approach of exploring subject of research (Jony and Serradell-López, 2018), developing conceptual or theoretical model (Jony *et al.*, 2025; Jony and Serradell-López, 2021). Systematic review research aims to provide valuable insights and identify avenues for further research in a critical domain (Jony and Serradell-López, 2019). The methodology used in this review is based on a hybrid of concept-context analysis and systematic literature review techniques, drawing inspiration from established works in the field (Jony and Hamim, 2024). This methodology ensures the review process's rigor, transparency, and scientific credibility, setting it apart from traditional narrative reviews. The methodology adopted in this study aims to rigorously examine the existing literature on cyber security attacks and their implications. This study adheres to the procedures as outlined in Jony *et al.* (2024) to organize the data. This review will go through three cycles: (1) preparing for the review, (2) performing the review, and (3) reporting the review (discussed in section 4, 5 and 6).

3.1. Planning the Review

The current study adopts a systematic literature review, employing an inductive reasoning approach to examine cybersecurity attacks comprehensively, focusing on AI-driven approaches. A specific set of well-defined criteria, such as search database, search keywords, and subject areas, is utilized to identify a corpus of relevant scholarly documents, facilitating a structured, integrated, and narrated literature review (Kraus *et al.*, 2022; Jony *et al.*, 2025). The criteria and data selection process outlined by Kraus *et al.* (2022) is followed to ensure a robust and comprehensive search. The primary databases used for this review include Web of Science (WoS), Google Scholar, Springer, ACM Digital Library, Scopus, and IEEE Xplore. In adherence to academic integrity principles, proper attribution and citation practices are followed throughout the study. The review process involves an unbiased and meticulous assessment of the identified literature, leading to the synthesis and interpretation of key findings related to AI-driven approaches for mitigating cybersecurity threats. Figure 2 outlines the step-by-step process of systematic reviewing and selecting relevant articles at each stage. By incorporating these additional databases and sources, this systematic review aims to provide an up-to-date analysis of AI-driven approaches in cybersecurity.

3.2. Performing the Review

During the selection process, our initial focus was on identifying scholarly publications that examine subjects such as "AI and cybersecurity", "AI-driven techniques", "malicious employment of AI", "AI-powered cyberattacks", "AI-based cyber-attacks", "in cybersecurity", and "Impact of AI in cybersecurity". The primary objective of this comprehensive search was to yield a substantial quantity of articles from many sources, then

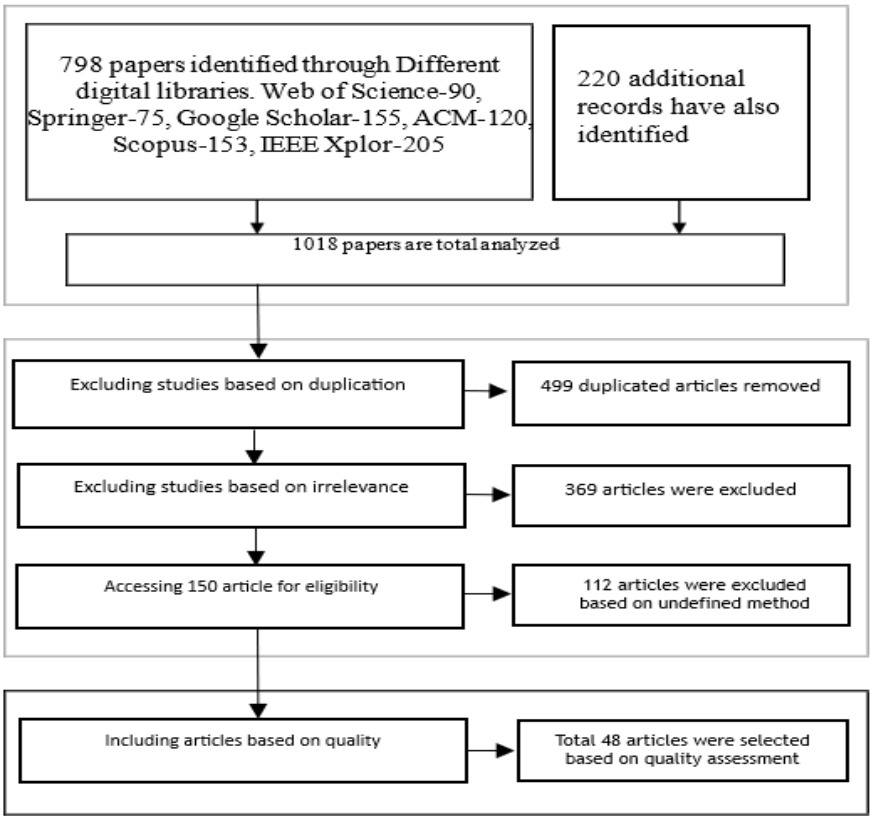


Figure 2: Step-by-Step Systematic Review Process

followed by a focused curation of scholarly journal articles sourced from esteemed academic publications in the field of information security and its associated subjects. The dataset underwent a comprehensive analysis of its titles, abstracts, keywords, and content, resulting in its ultimate curation. The selection process prioritized articles that offer valuable insights on the utilization of AI-driven techniques for the purpose of minimizing cybersecurity threats. The identification of patterns and trends in various attack vectors was accomplished by the categorization of selected articles based on the type of AI-driven assaults, the methodology employed, and the assessments of the resulting harm. Following this, pertinent data was synthesized and assessed in order to understand AI-driven methods for mitigating cybersecurity threats, including their characteristics and comprehensive countermeasures. The systematic approach employed in this technique facilitated the generation of more accurate conclusions and offered valuable insights for future research endeavors in the field of AI-driven cybersecurity. This paragraph aligns with the objectives of your evaluation since it incorporates the selected search terms and focuses on AI-based strategies for mitigating cybersecurity vulnerabilities.

Figure 2 is a flowchart depicting the step-by-step process of a systematic review, showing the stages of identification, screening, eligibility assessment, and inclusion of research papers for analysis.

- **Stage 1 (Extracting Information):** In the initial phase of the study, information was systematically extracted through a thorough search across nine electronic databases. The search yielded a total of 1018 article outlines, forming a substantial pool of potential articles for subsequent selection.
- **Stage 2 (Screening):** From the initially identified 1018 potential articles, as outlined in Table 2, it was observed that 499 papers were duplicates, appearing across multiple online resources. Subsequently, a screening process was implemented, focusing on the relevance of article titles. During this phase, 369 articles were excluded from consideration for this study due to their determined lack of suitability.
- **Stage 3 (Eligibility):** Acknowledging the significance and caliber of an article surpasses its abstract and title, comprehensive text-based selection criteria were employed to identify pertinent publications for this

study. A comprehensive evaluation was conducted on a corpus of 150 publications to determine their eligibility. Among these, 112 papers were excluded from further consideration due to deficiencies in their techniques or the presence of uncertainties.

- **Stage 4 (Inclusion):** The authors thoroughly assessed the remaining publications after conducting a quality assessment. Following the resolution of any selection issues of articles, a total of 48 primary papers emerged as the subject of this investigation's analysis.

4. Impacts of AI in Cybersecurity Landscape

AI is a comprehensive phrase that encompasses computational approaches that replicate human learning processes, allowing computers to explore and gain knowledge autonomously. Machine Learning (ML), an important aspect of AI is classified into supervised, unsupervised, and reinforcement learning. Unsupervised learning discovers patterns without the use of labelled data, whereas supervised learning uses labelled data to train models. To learn surroundings, reinforcement learning involves trial-and-error. Deep Learning (DL), a subclass of ML, leverages neural networks with several layers to learn sophisticated data representations automatically (Qiu *et al.*, 2016).

By reinventing service delivery and fortifying defences against developing cyber threats, AI is dramatically changing the cybersecurity landscape. Around 15% of businesses worldwide have adopted AI technical breakthroughs, realizing their potential influence (Babu *et al.*, 2012). While technological advancement has many advantages, it also raises security dangers. Businesses are recommended to continually upgrade AI security systems and apply stringent procedures to mitigate these dangers. AI helps by identifying potential threats through different coded signals and doing regular cybersecurity risk assessments. AI plays a critical role in preventing and responding to assaults such as viruses and phishing emails by automating security processes and improving detection capabilities, making cybersecurity more flexible and resilient in the face of sophisticated threats (Jony and Arnob, 2024). With continuously growing cyber-attacks and the rapid proliferation of electronics today, AI can assist hackers in being aware, automating attacks detection, and reacting more efficiently than normal automated and manual approaches (Zhuge, 2011).

AI is altering cybersecurity in a variety of ways. For starters, it improves efficiency by automating mundane processes, freeing up security analysts to focus on more complicated responsibilities such as incident response. Second, AI improves threat detection accuracy by outperforming traditional signature-based approaches in spotting nuanced threats such as new malware variants and delicate network traffic patterns (Arnob and Jony, 2024). Third, AI helps to reduce costs by automating tasks, reducing false positives, and speeding up issue response, resulting in significant operational savings. Furthermore, its real-time threat detection and response capabilities are critical in limiting potential damage from malicious activity as quickly as possible. Finally, AI's scalability processes large datasets efficiently, solving the challenges of dynamic cybersecurity landscapes. While AI improves scalability, human skill is still required for understanding results and making informed judgments. AI integration improves cybersecurity procedures, making them more adaptable, cost-effective, and resilient to emerging digital threats (Kaloudi and Li, 2020).

5. Applications of AI in Cybersecurity

The integration of AI has become a significant catalyst in the field of cybersecurity, among the constantly changing landscape of digital threats. With the continuous advancement of technology and the growing sophistication of cyber adversaries, it is evident that conventional security measures frequently fall short in adequately protecting valuable digital assets. AI came into being throughout the 20th century because of endeavours aimed at developing a system that operates autonomously, without the need for human cognitive interaction. The advancement served as a catalyst for additional investigation within the discipline prompting various endeavours to develop intelligent systems and robots that replicate human behaviour while mitigating their effects on human beings (Kaloudi and Li, 2020). AI has emerged as a valuable tool in enhancing cybersecurity measures by leveraging sophisticated algorithms and techniques including machine learning techniques, deep learning models, and adaptive functionalities. AI presents a diverse range of capabilities that can bolster the resilience of companies against cyber problems. These capabilities

encompass real-time threat identification and response, as well as the analysis of extensive datasets. This introduction examines the utilization of AI in many aspects of cybersecurity, highlighting its crucial function in reducing risks and actively responding to the changing landscape of digital security threats (Wiafe *et al.*, 2020).

5.1. Malware Detection

AI systems can be trained to identify and mitigate cyber risks and dangerous viruses, hence enhancing their efficacy in the field of cybersecurity. The proliferation of cyber security breaches has prompted the integration of AI into the realm of cybersecurity. The primary objective of the entire procedure is to optimize operational efficiency and enhance the level of accuracy (Chen and Liu, 2016). The presence of malware presents a substantial cybersecurity risk, and conventional antivirus software depends on signature-based detection that is constrained to recognized malware variations. The efficacy of this strategy can be compromised by the utilization of altered malware. On the other hand, AI-driven solutions utilize machine learning techniques to identify and analyse both familiar and unfamiliar security risks. These methods employ the analysis of extensive datasets, utilizing both labelled and unlabelled data for the purpose of training. AI utilizes methodologies such as static analysis, which involves the examination of file properties, and dynamic analysis, which entails the observation of file behaviour during execution. According to Faruk *et al.* (2021) AI-based malware detection presents a superior and efficient way in comparison to conventional antivirus software (Hossain Faruk *et al.*, 2021). This methodology can detect novel and unfamiliar malware variations that may evade detection through signature-based techniques. Nevertheless, AI possesses inherent limitations that prevent it from fully substituting human capabilities, mostly due to its programmed nature, which restricts its functionality to a predetermined set of tasks. Occasionally, the system fails to identify almost identical threats, leading to complications as it closely resembles genuine communication. The detection of risks by AI systems may be challenging because of the continuous evolution of cyber threats.

5.2. Phishing Detection

Organizations and individuals are the targets of phishing, a pervasive cyberthreat. Conventional detection techniques are limited in their efficacy to known assaults, as they depend on rules or blacklisting. Phishing attacks give rise to a multitude of hazards. Ansari *et al.* (2022) asserts that AI technologies effectively prevent most of these assaults. This has been linked to the fact that most organizations are concentrating on AI-based cyber security to safeguard their systems (Purkait, 2015). Conversely, AI-driven solutions analyze the content and structure of emails using machine learning algorithms that acquire knowledge from extensive datasets in order to detect patterns and irregularities that are suggestive of fraud. Furthermore, these solutions have the capability to evaluate user conduct by identifying suspicious actions such as responding to fraudulent emails or clicking on malicious links. This provides a more resilient and adaptable defence against ever-changing threats in comparison to rule-based methodologies (Ansari *et al.*, 2022).

5.3. Security Log Analysis

Security log analysis entails the examination of server logs with the purpose of identifying and resolving hardware or software failures, documenting alerts, warnings, and errors, and providing information that enables system administrators to comprehend the origins of potential system malfunctions. It functions as a preventive measure against malicious activities, including SQL injections, by facilitating the identification of pre-established attack methods (Cao *et al.*, 2017). Traditional security log analysis is limited in its ability to detect new threats, as it is dependent on rule-based systems. Powered by machine learning algorithms, AI-based security log analysis triumphs at the examination of massive amounts of log data in real time. Without known threat signatures, these algorithms can identify patterns and anomalies, allowing for prompt response to potential security intrusions (Jony and Arnob, 2024; Luo *et al.*, 2017). In addition, AI facilitates the detection of insider threats through the analysis of user behaviour across systems, which identifies anomalous occurrences like illicit access or out-of-the-ordinary data transfers. By adopting this proactive stance, organizations can proactively avert security incidents and data breaches, thereby establishing AI-powered security log analysis as a highly effective instrument for detecting and mitigating threats (Hieu *et al.*, 2019).

5.4. Dos/DDoS

DoS assaults are malevolent endeavors to impede or limit the regular operations of a specific server, service, or network, with the intention of preventing its intended users from accessing it. DDoS attacks have gained significant attention as a cyber-attack. To mitigate the increasing complexity of DDoS attacks, the integration of AI into cybersecurity protocols is vital. AI, in contrast to conventional static security protocols, is capable of dynamically comprehending and adjusting to shifting conditions. An AI-based approach to DDoS mitigation entails the training of algorithms to analyze network traffic patterns, identify anomalies that may serve as indicators of an impending attack, and provide real-time responses. AI systems possess the capability to detect established attack patterns and adjust to novel and developing threats by utilizing machine learning. This renders them exceptionally effective in mitigating the multifaceted and perpetually evolving characteristics of DDoS attacks (Singh, 2020; Zhang *et al.*, 2017; Saha *et al.*, 2022).

5.5. Man in the Middle

A cyber peril known as a Man-in-the-Middle (MitM) attack occurs when an unauthorized entity surveils and potentially modifies the communication exchanged between two entities without their awareness. MitM attacks possess the capability to perturb the regular data transmission between devices and cloud servers, thereby causing congestion and potentially jeopardizing the cloud environment's security and functionality (Conti *et al.*, 2016). To mitigate the impact of MitM assaults on cloud computing, the text suggests the integration of AI technology. By reducing the cognitive load on human operators and bolstering the security of the cloud environment, the AI system is intended to be more efficient. The framework can employ AI, more particularly a rule-based system utilizing IF-THEN logic, to intelligently detect anomalous behaviour in incoming data packets prior to their transmission to the cloud. The AI system, which has been programmed to distinguish between typical and atypical patterns, can detect deviations from anticipated conduct to identify potential MitM attacks (Ramiah *et al.*, 2019).

Table 2 shows different AI-powered cybersecurity methods, such as machine learning for finding malware and analyzing security logs, AI-powered phishing detection, AI algorithms for responding to DoS attacks, and rule-based systems to stop Man-in-the-Middle attacks. It focuses on finding and stopping cyber threats in real time that aren't normal.

Table 2: Overview of AI-Driven Cybersecurity Mitigation Techniques

Cybersecurity Threat	AI-Driven Approach	Description	References
Malware Detection	Machine Learning Techniques	AI systems use static and dynamic analysis to identify both known and unknown malware, surpassing conventional antivirus software in efficacy.	Chen and Liu (2016) and Hossain *et al.* (2021)
Security Log Analysis	Machine Learning Techniques	AI analyzes email content and structure to detect patterns indicative of phishing and evaluates user behavior to identify suspicious actions.	Purkait (2015) and Ansari *et al.* (2022)
Phishing Detection	Machine Learning Techniques	AI-powered analysis of security logs enables the identification of patterns and anomalies in real-time, aiding in the detection of insider threats and potential security intrusions.	Cao *et al.* (2017), Jony and Arnob (2024), Luo *et al.* (2017) and Hieu *et al.* (2019)
Denial-of-Service (DoS) Attacks	Network Traffic Analysis	AI algorithms are trained to identify anomalies in network traffic, providing real-time responses to both established and novel DDoS attacks.	Singh (2020), Zhang *et al.* (2017) and Saha *et al.* (2022)
Man-in-the-Middle (MitM) Attacks	Rule-Based Systems	AI utilizes IF-THEN logic to detect anomalous behavior in data packets, helping to prevent MitM attacks in cloud environments.	Conti *et al.* (2016) and Ramiah *et al.* (2019)

6. Machine Learning in Cyber Security

Despite the claims of numerous businesses, the book stresses that ML is the more common AI application in the field of cybersecurity. Cybersecurity has yet to fully adopt AI, despite the field's stated interest in giving robots human-like intelligence (Ivanichenko *et al.*, 2021). The focus is on ML, a branch of AI that makes use of statistical analysis, mathematical optimization, and data mining. Changing attack tactics, such as zero-day assaults, provides a problem for network security because conventional defenses have trouble keeping up in the lack of clear guidelines. ML steps in as a response, with the aim of real-time detection and isolation of malware. Quick action is taken to contain and prevent the propagation of dangerous code thanks to ML and AI-assisted judgments and isolation approaches. However, successfully identifying and responding to security events is a significant obstacle when applying AI or ML to network security, as deviations from patterns do not always signify incidents. In cybersecurity, it is essential to strike a balance between the necessity for precise detection and the desire to reduce false positives. ML algorithms in cybersecurity aim to let computers learn the optimal ways to fulfill their jobs without explicit programming, notably tackling the increasing issues of network security, where assaults frequently alter in appearance and vectors (Mohsin and Jony, 2024; Pinto *et al.*, 2023).

The basic goal of supervised ML in cybersecurity is to train machines to map inputs to outputs using real-world data that has been tagged for training purposes. This supervised method, which incorporates classification and regression issues, has proven effective in identifying malicious software and spam. The enormous databases holding malware files contribute to training these algorithms deeper, enhancing accuracy and lowering false positives and negatives. Recurrent Neural Network (RNN)-based classification algorithms are used to differentiate between good and bad actors, demonstrating the power of supervised ML in strengthening cybersecurity defenses (Ferdous *et al.*, 2024).

Unsupervised machine learning involves computers independently identifying patterns and relationships in data without relying on labeled samples. This methodology incorporates methodologies such as cluster analysis and dimension reduction, which are valuable for detecting atypical network activity and examining extensive data sets. Unsupervised machine learning is becoming increasingly popular in the field of cybersecurity since it has the potential to improve pattern detection by overcoming the constraints of labeled data (Pinto *et al.*, 2023).

7. Discussion

In the context of the rapidly evolving field of cybersecurity, the integration of AI has emerged as a robust defense mechanism against the increasing threat of cyber-attacks. Considering the significant advancements observed in the domain of Network Security in recent times, particularly in relation to the incorporation of, it is advisable to take caution while considering the scope of its prospective implementations. There exists a tendency to erroneously hold the belief that AI possesses the ability to serve as a universal solution for all cyber security challenges, or to blindly expect its triumph in scenarios where traditional approaches have failed (Arnob and Jony, 2024).

This systematic review explores the intersection of AI and cybersecurity, investigating its various uses, assessing its effectiveness, and identifying the persistent issues that exist in this domain. The present escalation in cyber threats, propelled by the advanced nature of AI-enabled attacks, necessitates the implementation of a proactive and adaptable protection plan. Conventional cybersecurity techniques, originally developed to counter traditional attacks, are becoming progressively insufficient when confronted with the emergence of AI-powered cyber weapons (Adi *et al.*, 2022). The advent of 'offensive AI' has empowered cybercriminals to carry out focused attacks with unparalleled velocity and intricacy, surpassing the capabilities of human-operated systems (Kamoun *et al.*, 2020). This transition marks the onset of a novel period in which cyber threats exhibit heightened intelligence and possess the capacity to independently adjust and make decisions in response to their specific target environment. The dynamic strategies employed by hackers emphasize the critical need to implement sophisticated AI-based methods in order to effectively address and reduce cybersecurity vulnerabilities. Contemporary attackers exploit AI in order to execute social engineering tactics with unprecedented degrees of effectiveness, beyond previously perceived limits (Salahdine and Kaabouch,

2019). These attacks possess the ability to acquire knowledge from contextual data and demonstrate adaptability by imitating reliable aspects of the digital realm or attacking flaws with remarkable accuracy. The issues presented by cyber threats caused by transcend beyond the realm of technological considerations. The escalating intricacy and magnitude of these attacks provide a significant challenge for cybersecurity teams, as they encounter difficulties in locating proficient experts who possess the necessary skills to effectively mitigate these advancing dangers. As enterprises undergo the process of adopting more secure digital environments, such as cloud storage services, they continue to face vulnerabilities, hence requiring a fundamental change in cybersecurity approaches.

The analysis demonstrates that AI-powered attacks employ intricate obfuscation methods, often modifying discernible attributes to such an extent that conventional antivirus software faces challenges in detecting them. The deliberate malicious utilization of, as demonstrated by elusive attacks, underscores the necessity for inventive strategies capable of effectively addressing the flexibility and complexity of AI-based risks (Hu *et al.*, 2021). Despite the numerous hurdles, AI-driven cybersecurity methods show potential in strengthening defenses against the constant growth of cyber threats. Through the examination of extensive datasets in real-time, AI algorithms possess the capability to identify anomalies and trends that may signify possible breaches. This ability enables the establishment of a security framework that is adaptable and responsive. Nevertheless, given the escalating potential of AI as a tool for individuals with bad intent, it is crucial to prioritize ongoing research and innovation in order to maintain a competitive advantage. Figure 3 tracks the progression of AI's role in cybersecurity across five stages:

- **Stage 1 – Initial:** AI starts being recognized for its potential in cybersecurity.
- **Stage 2 – Development:** AI capabilities are developed to a point where proactive defences become possible.
- **Stage 3 – Maturity:** AI systems start making autonomous decisions and offensive AI begins to appear.
- **Stage 4 – Advanced:** The utilization of AI by malicious actors for the purpose of executing intricate social engineering and adaptive attacks has been observed. Additionally, there is a recognition of the existence of a technical skills gap.
- **Stage 5 – Future Outlook:** The focus is on continuous innovation to stay ahead of AI-powered attacks and challenges.

The checkmarks (✓) indicate the presence of a feature or capability, while the crosses (X) indicate its absence or underdevelopment at that stage (Figure 3).

AI Development Stages in Cybersecurity	Stage 1: Initial	Stage 2: Development	Stage 3: Maturity	Stage 4: Advanced	Stage 5: Future Outlook
AI Integration in Cybersecurity	✓	✓	✓	✓	✓
Proactive Defense Capabilities	X	✓	✓	✓	✓
Autonomous Decision-Making	X	X	✓	✓	✓
Offensive AI Threats	X	X	✓	✓	✓
`Advanced Social Engineering	X	X	X	✓	✓
Adaptive Attack Strategies	X	✓	✓	✓	✓
Technical Skill Gap	✓	✓	✓	✓	✓
Need for Continuous Innovation	✓	✓	✓	✓	✓
Real-Time Anomaly Detection	X	✓	✓	✓	✓
AI-Powered Obfuscation Challenges	X	X	✓	✓	✓

Figure 3: Evolution of AI Capabilities in Cybersecurity: A Stage-by-Stage Matrix

8. Conclusion

The exploration of AI in cybersecurity presented in this review underlines its dual role as both a tool for attackers and a countermeasure. While AI-driven methods have advanced the field, they also present new challenges that demand a dynamic and innovative approach to security. The matrix of AI capabilities in cybersecurity delineates a clear evolution from initial recognition to an outlook that emphasizes continuous innovation. The necessity for adaptive defense strategies, real-time anomaly detection, and the bridging of technical skill gaps has never been more pronounced. As we stand at the cusp of a new era in cyber threats, it is paramount for cybersecurity entities to leverage AI's full potential while remaining vigilant of its use in malicious hands. Embracing this dual-edged sword requires a delicate balance of embracing technological advances and cultivating a robust, forward-looking research ecosystem to safeguard digital assets against the ever-advancing landscape of cyber threats.

9. References

Adi, E., Baig, Z. and Zeadally, S. (2022). Artificial Intelligence for Cybersecurity: Offensive Tactics, Mitigation Techniques and Future Directions. *Applied Cybersecurity & Internet Governance*, 1(1), 1-23. doi: 10.5604/01.3001.0016.0800.

Ansari, M.F., Sharma, P.K. and Dash, B. (2022). Prevention of Phishing Attacks Using AI-Based Cybersecurity Awareness Training. *International Journal of Smart Sensor and Adhoc Network*, March, 61-72. doi: 10.47893/ijssan.2022.1221.

Arnob, A.K.B. and Jony, A.I. (2024). Enhancing IoT Security: A Deep Learning Approach with Feedforward Neural Network for Detecting Cyber Attacks in IoT. *Malaysian Journal of Science and Advanced Technology*, 4(4), 413-420.

Babu, G.D., Anandakuma, N.N. and Muralidhar, D. (2012). Countermeasures against DPA Attacks on FPGA Implementation ofAES. *Journal of Artificial Intelligence*, 5(4), 186-192. doi: 10.3923/jai.2012.186.192.

Biswas, T., Ferdous, F.S. and Jony, A.I. (2024). Binary and Multi-Class Prediction of DDoS Attack Using Deep Learning Models. *International Journal of Data Science and Big Data Analytics*, 4(2), 49-58.

Cao, Q., Qiao, Y. and Lyu, Z. (2017). Machine Learning to Detect Anomalies in Web Log Analysis. in *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*.

Chen, Z. and Liu, B. (2016). Lifelong Machine Learning. *Synthesis Lectures on and Machine Learning*, 10(3), 1-145. doi: 10.2200/s00737ed1v01y201610aim033.

Conti, M., Dragoni, N. and Lesyk, V. (2016). A Survey of Man in the Middle Attacks. *IEEE Communications Surveys & Tutorials*, 18(3), 2027-2051. doi: 10.1109/comst.2016.2548426.

Dixon, W. and Eagan, N. (2019). AI will Power a New Set of Tools and Threats for the Cybercriminals of the Future.

Ferdous, F.S., Biswas, T. and Jony, A.I. (2024). Enhancing Cybersecurity: Machine Learning Approaches for Predicting DDoS Attack. *Malaysian Journal of Science and Advanced Technology*, 4(3), 249-255.

Guembe, B., Azeta, A., Misra, S., Osamor, V.C., Fernandez-Sanz, L. and Pospelova, V. (2022). The Emerging Threat of AI-Driven Cyber Attacks: A Review. *Applied*, 36(1), March. doi: 10.1080/08839514.2022.2037254.

Hamadah, S. and Aqel, D. (2020). Cybersecurity Becomes Smart Using Artificial Intelligent and Machine Learning Approaches: An Overview. *ICIC Express Letters, Part B: Applications*, 11(12), 1115-1123.

Hieu Le Ngoc, Tran Cong Hung, Nguyen Đuc Huy and Hang, T. (2019). Early Phase Warning Solution about System Security Based on Log Analysis. *2019 6th NAFOSTED Conference on Information and Computer Science (NICS)*, December. doi: https://doi.org/10.1109/nics48868.2019.9023899.

Hossain Faruk, M.J. *et al.* (2021). Malware Detection and Prevention Using Techniques. *IEEE Xplore*, December 01. https://ieeexplore.ieee.org/document/9671434/

Hu, Y. *et al.* (2021). Security: Threats and Countermeasures. *ACM Computing Surveys*, 55(1), 1-36. doi: 10.1145/3487890.

Ivanichenko, Y., Sablina, M. and Kravchuk, K. (2021). Use of Machine Learning in Cyber Security. *Cybersecurity: Education, Science, Technique*, 4(12), 132-142. doi: 10.28925/2663-4023.2021.12.132142.

Jony, A.I. and Arnob, A.K.B. (2024). A Long Short-Term Memory Based Approach for Detecting Cyber Attacks in IoT Using CIC-IoT2023 Dataset. *Journal of Edge Computing*, 3(1), 28-42. Available: https://doi.org/10.55056/jec.648

Jony, A.I. and Arnob, A.K.B. (2024). Securing the Internet of Things: Evaluating Machine Learning Algorithms for Detecting IoT Cyberattacks Using CIC-IoT2023 Dataset. *International Journal of Information Technology and Computer Science*, 16(4), 56-65.

Jony, A.I. and Hamim, S.A. (2023). Navigating the Cyber Threat Landscape: A Comprehensive Analysis of Attacks and Security in the Digital Age. *Journal of Information Technology and Cyber Security*, 1(2), 53-67.

Jony, A.I. and, Hamim, S.A. (2024). Empowering Virtual Collaboration: Harnessing AI for Enhanced Teamwork in Higher Education. *Educational Technology Quarterly*, 2024(3), 337-359.

Jony, A.I. and Serradell-López, E. (2018). Key Performance Indicators of Virtual Teamwork: A Conceptual Framework. *ICERI2018 Proceedings*, 5059-5068, IATED.

Jony, A.I. and Serradell-López, E. (2019). Effective Virtual Teamwork Development in Higher Education: A Systematic Literature Review. *Edulearn19 Proceedings*, 873-882. doi: 10.21125/edulearn.2019.0285.

Jony, A.I. and Serradell-López, E. (2021). Key Factors that Boost the Effectiveness of Virtual Teamwork in Online Higher Education. in Visvizi, A., Lytras, M.D. and Aljohani, N.R. (Eds.), *Research and Innovation Forum 2020 (RIIFORUM2020)*, Springer Proceedings in Complexity. https://doi.org/10.1007/978-3-030-62066-0_15

Jony, A.I., Hamim, S.A. and Serradell-López, E. (2025). AI-Powered Teamwork in Higher Education: A Comprehensive Systematic Review of Effectiveness and Challenges. in *Multidisciplinary Movements in AI and Generative AI*, 65-97.

Jony, A.I., Lara Navarra, P. and Serradell-López, E. (2025). Key Functionalities of Personalized AI Tutoring Systems that Boost Learning Outcomes in Higher Education. in *ICERI2025 Proceedings*, 8200-8209.

Jony, A.I., Rithin, A.T. and Edrish, S.I. (2024). A Comparative Study and Analysis of Text Summarization Methods. *Malaysian Journal of Science and Advanced Technology*, 4(2), 118-129.

Kaloudi, N. and Li, J. (2020). The AI-Based Cyber Threat Landscape. *ACM Computing Surveys*, 53(1), 1-34. doi: 10.1145/3372823.

Kamoun, F., Iqbal, F., Esseghir, M.A. and Baker, T. (2020). AI and Machine Learning: A Mixed Blessing for Cybersecurity. *IEEE Xplore*, October 01. https://ieeexplore.ieee.org/document/9297323

Kraus, S. *et al.* (2022). Literature Reviews as Independent Studies: Guidelines for Academic Practice. *Review of Managerial Science*, 16(8), 2577-2595. doi: 10.1007/s11846-022-00588-8.

Kraus, S., Durst, S., Ferreira, J.J., Veiga, P., Kailer, N. and Weinmann, A. (2022). Digital Transformation in Business and Management Research: An Overview of the Current Status Quo. *International Journal of Information Management*, 63, April, 102466. doi: 10.1016/j.ijinfomgt.2021.102466.

Luo, X., Wang, J., Shen, Q., Wang, J. and Qi, Q. (2017). User Behavior Analysis Based on User Interest by Web Log Mining. *IEEE Xplore*, November 01. https://ieeexplore.ieee.org/abstract/document/8215435/ (accessed Feb. 01, 2023).

Mohsin, M. and Jony, A.I. (2024). A Comparative Analysis of Medical IoT Device Attacks Using Machine Learning Models. *Malaysian Journal of Science and Advanced Technology*, 4(4), 429-439.

Pinto, S.J., Siano, P. and Parente, M. (2023). Review of Cybersecurity Analysis in Smart Distribution Systems and Future Directions for Using Unsupervised Learning Methods for Cyber Detection. *Energies*, 16(4), 1651.

Purkait, S. (2015). Examining the Effectiveness of Phishing Filters against DNS Based Phishing Attacks. *Information & Computer Security*, 23(3), 333-346. doi: 10.1108/ics-02-2013-0009.

Qiu, J., Wu, Q., Ding, G., Xu, Y. and Feng, S. (2016). A Survey of Machine Learning for Big Data Processing. *EURASIP Journal on Advances in Signal Processing*, 2016(1). doi: 10.1186/s13634-016-0355-x.

Ramiah Chowdary, P., Challa, Y. and Jitendra, M.S.N.V. (2019). Identification of MITM Attack by Utilizing Mechanism in Cloud Environments. *Journal of Physics: Conference Series*, 1228, 012044. doi: https://doi.org/10.1088/1742-6596/1228/1/012044.

Saha, S., Priyoti, A.T., Sharma, A. and Haque, A. (2022). Towards an Optimized Ensemble Feature Selection for DDoS Detection Using Both Supervised and Unsupervised Method. *Sensors*, 22(23), 9144. doi: 10.3390/s22239144.

Salahdine, F. and Kaabouch, N. (2019). Social Engineering Attacks: A Survey. *Future Internet*, 11(4), 89. doi: https://doi.org/1052.3390/fi11040089.

Singh, J. (2020). Mitigating DoS and DDoS Based Attacks: An Approach. *International Journal of Innovative Science and Research Technology*, 5(5).

Thanh, C.T. and Zelinka, I. (2019). A Survey on in Malware as Next-Generation Threats. *Mendel*, 25(2), 27-34. doi: 10.13164/mendel.2019.2.027.

Usman, M., Jan, M.A., He, X. and Chen, J. (2019). A Survey on Representation Learning Efforts in Cybersecurity Domain. *ACM Computing Surveys*, 52(6), 1-28. doi: 10.1145/3331174.

Usman, M., Jan, M.A., He, X. and Chen, J. (2019). A Survey on Representation Learning Efforts in Cybersecurity Domain. *ACM Computing Surveys*, 52(6), 1-28. doi: 10.1145/3331174.

Wiafe, I., Koranteng, F.N., Obeng, E.N., Assyne, N., Wiafe, A. and Gulliver, S.R. (2020). Artificial Intelligence for Cybersecurity: A Systematic Mapping of Literature. *IEEE Access*, 8, 146598-146612. doi: 10.1109/access.2020.3013145.

Zhang, B., Zhang, T. and Yu, Z. (2017). DDoS Detection and Prevention Based on Techniques. *IEEE Xplore*, December 01. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8322748

Zhuge, H. (2011). Semantic Linking Through Spaces for Cyber-Physical-Socio Intelligence: A Methodology. 175(5-6), 988-1019. doi: 10.1016/j.artint.2010.09.009.

Zouave, E., Bruce, M., Colde, K., Jaitner, M., Rodhe, I. and Gustafsson, T. (2020). Artificially Intelligent Cyberattacks. Swedish Defence Research Agency, FOI, Tech. Rep. FOI.

Cite this article as: Sultanul Arifeen Hamim and Akinul Islam Jony (2026). A Systematic Review of AI-Driven Approaches for Mitigating Cybersecurity. *International Journal of Data Science and Big Data Analytics*, 6(1), 55-67. doi: 10.67191/IJDSBDA.6.1.2026.55-67.