Journal of
Cyber Security

Tech Science Press

DOI: 10.32604/jcs.2024.056164

REVIEW

Enhancing Cyber Security through Artificial Intelligence and Machine Learning: A Literature Review

Carlos Merlano[*]

Purdue Polytechnic Institute, Purdue University, West Lafayette, IN 47907, USA

*Corresponding Author: Carlos Merlano. Email: cmerlano@purdue.edu

Received: 15 July 2024 Accepted: 10 October 2024 Published: 06 December 2024

ABSTRACT

The constantly increasing degree and frequency of cyber threats require the emergence of flexible and intelligent approaches to systems' protection. Despite the calls for the use of artificial intelligence (AI) and machine learning (ML) in strengthening cyber security, there needs to be more literature on an integrated view of the application areas, open issues or trends in AI and ML for cyber security. Based on 90 studies, in the following literature review, the author categorizes and systematically analyzes the current research field to fill this gap. The review evidences that, in contrast to rigid rule-based systems that are static and specific to a given type of threat, AI and ML are more portable and effective in large-scale anomaly detection, malware classification, and prevention of phishing attacks by analyzing the data, learning the patterns, and improving the performance based on new data. Further, the study outlines significant themes, such as data quality, integration, and bias with AI/ML models, and underscores overcoming barriers to undertaking standard AI/ML integration. The contributions of this work are as follows: a thorough description of AI/ML applications in cyber security, discussions on the critical issues, and relevant opportunities and suggestions for future research. Consequently, the work contributes to establishing directions for creating and implementing AI/ML-based cyber security with demonstrable returns of technical solutions, organizational change, and ethicist interventions.

KEYWORDS

Artificial intelligence; machine learning; cyber security; threat detection; vulnerability assessment; network security; security automation; adversarial machine learning; explainable AI

1 Introduction

This paper discusses the role of artificial intelligence (AI) and machine learning (ML) in enhancing cyber security and their ability to adapt to dynamic and ever-changing cyber threats. Doing so has transformative potential for threat detection and response mechanisms in cyber security.

An ever-accelerating pace of market digitalization shapes the client-orientation process in all sectors, leaving people with enormous opportunities to be constantly connected and thoroughly satisfied. Nonetheless, this has also seen the rise in—and the increased frequency and complexity of—cyber threats in the digital age [1–3]. The contemporary cyber security environment constantly faces threats and attacks targeting organizations and individuals, including professional ransomware attacks, state-sponsored cyber espionage, amateur phishing attempts, and large-scale malware intrusions [4,5]. The cost of these incidents is rather worrying [6], with billions of US dollars being lost yearly on account of loss of data, revenue generation loss due to system crashes, and theft of intellectual property [6]. The social impact is also enormous; people and organizations have a loss of reputation, privacy invasion, and even physical harm because of cyber attacks [2]. The weaknesses of the standard for defined security policies rely on such evolving threats [7,8].

On the one hand, it is the volume and velocity of the threats; on the other hand, it is the adaptability and evolvability of the threats becoming more and more sophisticated [9] that make the traditional methods of identifying threats based on their signatures and trying to erect protective perimeters almost useless [10]. Many new threats, including zero-day vulnerabilities, advanced persistent threats (APTs), and polymorphic malware, have emerged to compound the weakness of traditional security techniques [11,12]. The constant emergence of new and more advanced threats makes it increasingly necessary to find new ways for identifying and preventing them before they happen to the detriment of some helpless victim. For instance, Atiku et al. [1] demonstrated that AI and ML technology can address these challenges by enabling the development of intelligent systems that can analyze large data sets for real-time detection and response to threats. To illustrate, Xu et al. [13] found that computer programs can detect any abnormality in network traffic, user behavior or system logs that may indicate an intrusion.

This literature review is motivated by the critical necessity of analyzing how AI and ML can enhance cyber security. The growth rate of threats in the cyber realm and the limitations of conventional thinking about addressing them point toward the need to look for new approaches that will be effective in the network enabled capability (NEC) context. AI and ML can process large amounts of data, extract patterns, and learn from experience, which are promising areas for improving threat identification, mitigation, and prevention measures. However, to successfully integrate AI and ML into cyber security practices, it is crucial to understand how they work, what problems they can solve, and what pending and prospective opportunities are available. This literature review will provide such an understanding and be a valuable resource for cyber security researchers, practitioners, and decision-makers. This research is essential because it will lead to more robust and resilient cyber infrastructures. Organizations can employ AI and ML to strengthen their defense mechanisms and provide them with predictive capabilities to manage threats proactively. Data can be continuously monitored through AI-based systems and analyzed to identify possible points of vulnerability and predict attack vectors. Hence, organizations can take pre-emptive measures against risk [14]. Nevertheless, Giudice et al. [15] underscored the importance of organizational acceptance of AI and ML technologies.

Further, Buczak et al. [14] contended that AI and ML can be leveraged to automate duties such as threat hunting, incident response, and vulnerability assessment, hence allowing human analysts to focus on strategic activities. Automation improves cyber security operations' effectiveness, especially with the scarcity of skilled personnel [9].

The article investigates the application of AI and ML in various aspects of cyber security infrastructure, including network, cloud, and application security. It also probes additional issues, including adversarial attacks, the need for explainable AI (XAI), and ethical concerns about AI use in security scenarios. This research addresses these intricate matters by providing a comprehensive balance sheet on shaping future cyber security through AI and ML. The central question of this study is how organizations effectively integrate AI and ML into a cyber security framework while addressing limitations/challenges.

The article has six sections: Introduction (Section 1), Literature Review (Section 2), Methodology (Section 3), Results and Discussion (Section 4), Future Research Directions (Section 5), and Conclusion (Section 6).

2 Literature Review

This literature review provides a comprehensive and up-to-date overview of state-of-the-art AI and ML for cyber security, identifies critical challenges and opportunities, and informs future research directions in this rapidly evolving field.

While this review focuses on the period from 2015 onwards, which paved the way for the current development of AI and ML for cyber security, many concepts of AI and ML in cyber security originated several decades ago, and specific papers shaped the field. These early studies focused on enabling different AI and ML approaches to deal with cyber security issues, including intrusion detection, anomaly detection, and malware.

It is compelling to see the continuous evolution of models in intrusion detection theory. One of the earliest models, developed by Denning [16] in 1987, laid the theoretical basis for all Intrusion Detection Systems (IDS) created over the years. Most of these models, including the early one, draw on AI and ML to enhance their performance, showcasing the field's intriguing progress.

Another superlative work of the old generation is done by Forrest et al. [17] on anomaly detection for Unix processes. This research suggested that the potential of AI/ML in discovering new forms of threats is immense. An intrusion detection model can be developed based on departing from the system's normal behavior, showcasing the field's capabilities. A related work [18] proposed employing autonomous agents for intrusion detection, which led to the formulation of intelligent and adaptive security systems.

In another work, Lee et al. [19] examined data mining and ML to construct intrusion detection models that can uncover important information from massive security data. Mukkamala et al.'s [20] comparative evaluation of artificial neural networks and support vector machines in intrusion detection supported the application of ML methods to improve protection schemes.

Drawing on the above AI/ML in cyber security history, in the following sections of the paper, the author will analyze present-day security threats and how AI and ML need to be further developed and incorporated into the current cyber security environment.

2.1 Current Cyber Security Challenges

2.1.1 Overview of Current Threats

Threats in the cyber domain have significantly transformed in the last few years from an increase in sheer numbers and, quite worryingly, making them hard to counter. New studies stress AI and ML usage in emerging threats, stating their capability to boost IDS [21–23] and capitalize on cyber threat intelligence platforms [24]. Malware, phishing, and ransomware attacks are highly effective in moving beyond traditional protection tools' reach, as Husák et al. [4] and Sanghani et al. [25] suggested. According to Xiao et al. [11] and Shu et al. [7], zero-day vulnerabilities are software flaws the vendor has not identified, yet, and no patch exists. Consequently, AI technologies surge as crucial technological tools in forming predictive offense measures that can preemptively tackle new threats [12].

Wagner et al. [5] mentioned that state-sponsored attacks and APT intensify cyber security. For instance, state actors leverage enormous efforts and sophisticated instruments for spying, damaging, and misleading concerning critical infrastructure, government departments, and corporate entities. APTs are quiet and sneak into a system extensively, loitering in it, shifting their tactics, siphoning off data or subverting it [26]. These threats are usually sophisticated and often use spear phishing or watering hole attacks that are not easily recognizable.

Also, the advancement of various digital systems and Internet of things (IoT) devices has increased the attack vectors, thus exposing potential vulnerabilities. As highlighted by [27], IoT devices are primarily unsafe and do not have adequate security features; hence, they can be easily manipulated as a gateway into further networks and infect other networks in a domino effect. Cyber threats' immense and dynamic nature requires a change in perception from reactive to preventive, innovative, and an approach involving AI and ML [2,3]. Further, Industry 4.0 and the growing level of interconnectivity have enlarged these threats, which is why new methods based on AI need to be used to protect networks from constant threats [10].

Yet, AI systems are not safe from cyber threats and are prone to attacks. Adversaries can circumvent AI-based defenses, attack their algorithms, and attempt to tamper with the training datasets to render AI security tools ineffective [28].

Due to the constantly evolving hostile environment, organizations must reevaluate conventional security methods that rely solely on signature detection and perimeter defenses. Modern cyber threats are dynamic and adaptive, so it is high time to shift to proactive, intelligence-based, and 'learning' security measures with the help of AI and ML.

2.1.2 Limitations of Traditional Cyber Security Approaches

Traditional cyber security measures predominantly rely on signature detection technology, which implies comparing the signatures of input data against the database of known malicious signatures (e.g., virus definitions, IDS signatures). The traditional cyber security strategy works well against known threats but is limited to novel attacks, as References [8,7] argued that such attacks do not possess pre-existing signatures.

Another threat to organizations is zero-day attacks, in which the attackers successfully locate a new and unrecognized form of a cyber risk that has yet to have a countermeasure on the market. These attacks have the disadvantage of being able to avoid detection by the system defenses because they do not resemble any other pattern and leave an organization open to highly advanced, targeted attacks with devastating results [12].

Static defenses, including firewalls and intrusion prevention systems, cannot effectively counter cyber adversaries' modern strategies and tendencies. Malware authors always devise ways to bypass security measures; thus, rules and configurations become outdated quickly. The emerging threats require novelties and patches frequently, which are complex and could bring new threats into operation [10]. This circumstance explains why there is a need for advanced proactive security that employs AI and ML, which are self-learning and can counter new types of attacks [29].

Furthermore, typical anti-virus strategies are oriented toward outside threats, while insider threats are usually neglected. However, this assumption is no longer valid in today's environment, where insider threats and lateral movement within the network are significant concerns [6,30]. In other cases, standard approaches to recognizing or stopping adversarial actions from a penetrated internal host or user have weaknesses.

Countering modern threats will require more than the typical cyber security approaches. A more aggressive approach, greater sensitivity to new threats, and more significant innovation are now necessary in the fight against the threats of the new century, which is why AI and ML play a significant role.

2.2 Fundamentals of AI and ML in Cyber Security

2.2.1 Introduction to AI and ML

AI is a way in which artificial machines imitate the ability of human thinking by being trained and developed to think and learn like humans. AI is a broad field comprising ML, which involves using algorithms to learn from data and make decisions. This is why the learning process lies at the heart of the application of ML in cyber security: discovering new and unknown threats, which a system based on a set of rules can miss [31,32]. In cyber security, these technologies analyze large amounts of data to determine signs representing a threat. The modern world requires inserting AI and ML techniques in cyber security to improve performance [8]. The threats are progressive, and sophisticated solutions are needed to overcome them. For instance, several authors highlight the prospect of applying AI technology, such as ML and deep learning, to enhance cyber security [2,33]. Enhancing AI features enables cyber security specialists to prevent and combat threats more efficiently.

AI and related digital technologies are instrumental in implementing security solutions and data analysis to prevent systems from being vulnerable to new threats [9]. AI and ML have been adopted in industry and research to solve emerging and critical problems, focusing on cyber security as the vital area to address broad-based malware threats in critical systems [10,34]. Scholars have identified these technologies' primary roles in the different aspects of cyber security [35,36]. As stated by Ahsan et al. [37], these technologies are more scalable and practical in identifying malware than traditional techniques that require more focus and intervention from human experts.

AI efficiently interprets various cyber security data, carrying out activities such as asset categorization, control assignment, vulnerability handling, and threat identification, among others [38]. To address successive anomalous behaviors, cyber security specialists can use AI and ML and work preventively in the fight against cyber threats [39].

AI and ML significantly improve the ability of IoT to describe cyber behavior, analyze risks and vulnerabilities, and identify cyber attacks [40]. According to [41], advanced analytical techniques such as ML, data mining, deep learning, and expert systems are essential for strengthening security. The number of threats is rising, and cyber security operators have turned to ML to improve defense efficiency [42].

Supervised and unsupervised ML techniques, which involve labeling data and finding hidden patterns, respectively, possess the potential to provide solutions to cyber security issues, such as intrusion detection systems [43]. In addition, AI and ML algorithms are used frequently in practicing cyber security, which shows their importance in enhancing the security system [44]. These technologies permit the work on emerging patterns of security incidents out of the cyber security dataset and allow the creation of intelligent and automated systems based on the outcomes [9].

Conclusively, the narratives indicated that integrating AI and ML technologies is central to enhancing cyber security solutions. These technologies, when adopted, enable cyber security specialists to improve threat identification, apply machine-based solutions to several problems, and provide better security against new types of threats. Combining AI and cyber security is one prominent emerging field in protecting data and sensitive structures. In addition, ongoing research on adversarial ML, such as the one by Ijiga et al. [45], holds promise for enhancing threat detection and prevention schemes.

2.2.2 Applications in Cyber Security

AI and ML are valuable for cyber security by providing a more effective and proactive approach to threat detection and response [22]. These applications are virtually in all areas, and one of the most familiar is detecting threats. Thanks to ML algorithms, AI systems can process big data from various sources, such as network traffic, system logs, and users' behaviors, and realize patterns that are out of normal parameters [14]. This capability makes it possible to identify new threats that the other detection methods largely fail to recognize, such as zero-day threats.

Other learning algorithms include clustering and anomaly detection, where, based on traffic patterns, the system can spot abnormal traffic and label it as malicious traffic even without the usual signature [13]. However, with supervised learning techniques, one has training sets of preferably labeled malicious and benign traffic to build high-precision models.

Moreover, AI and ML are also commonly used in network security management. Other typical network management practices include structured techniques, which presuppose manual configuration and monitoring. Conversely, AI can self-regulate and supervise network activities, dynamically learn, and detect risks that harm the network [46]. The specific techniques and advancements in AI/ML include anomaly detection as the most basic approach in cyber security, centered on detecting squares familiar to the system, but quite different from the everyday norm. In anomaly detection, some conventional techniques involve rule-based systems or statistical methods, which sometimes are helpless in figuring out an innovative approach to intrusion. Newer techniques in AI/ML have made it possible to build significantly enhanced anomaly detection systems capable of processing vast amounts of data and further analyzing for amplified patterns in the observed transactions that could indicate abnormally performed actions [14,38]. These systems utilize methods such as unsupervised learning, clustering, and deep learning to design models capable of detecting between normal and abnormal traffic without the benefit of learning from previous samples or data. Research utilizing AI/ML for anomaly detection has yielded good outcomes in cyber security areas such as intrusion detection, malware analysis, and fraud detection [14,38,43].

Furthermore, when it comes to automated malware analysis, deep learning models, including convolutional neural networks and recurrent neural networks, have drawn much interest and have been proven effective due to the models' potential to identify features and patterns directly from raw data [32,37]. In addition, intrusion detection and prevention systems (IDPS) are essential in countering threat attempts at networks and systems' invasion and use. Intelligent attackers can easily circumvent the traditional IDPS, which depends on rules or signatures. AI/ML incorporated IDS that can adapt to the interpretations of the data traffic and dynamics of the users to establish unsuitable changes and intrusions as part of the real-time data analysis. The advantage of these systems is that they can learn from new attack patterns and techniques and offer more preventive and robust protection. Researchers have applied AI/ML to IDPS, exploring several ML algorithms and deep learning models to enhance system performance [21,23]. AI/ML is also helpful in identifying written content such as emails, websites, and other forms of communication that may contain phishing attempts. Besides, the features of these communications–the text, the links, and others–can be processed using natural language processing (NLP) and ML to help reveal phishing indicators such as phony message content when the domains used are fakes or the attachments are malicious [25,37].

Lastly, ML and AI approaches can be applied to parse the logs of activities performed by the users, network traffic, or any other data that would help derive the expected behavior of the network. These can then be classified as some security anomaly–perhaps an insider threat or a compromised account, among others. Incorporating AI/ML into user and entity behavior analytics [30] is proving to be highly effective in identifying new and improved attack patterns that often go unnoticed by conventional security approaches [14]. This effectiveness of AI/ML in identifying new and improved attack patterns instills confidence in the systems' capabilities. AI/ML extends beyond user behavior; for instance, AI-guided network security can range from identifying and preventing security threats in networks to alert prioritization and suggesting modifications to the network's configuration. Such systems can come up with conclusions from previous experiences and incidents, making it easier to develop network protective measures and offering a more robust way of protecting the network from threats. Additionally, security information and event management systems benefit from AI/ML algorithms as these can correlate events from different sources and identify new contexts that were previously unknown, completing and making this information actionable more quickly and accurately than a human analyst [47].

AI and ML help revolutionize cyber security by performing repetitive tasks, bolstering the defenses to stop threat actors, and helping cyber security specialists maintain their advantage in ongoing warfare.

2.3 Advantages of Integrating AI into Cyber Security

2.3.1 Enhanced Threat Detection

Another important feature of AI technologies is their ability to learn and identify features and deviations within big data, which is critically important in cyber security. Also, their ability results in better identification of advanced cyber threats that often go unnoticed by other security solutions. AI tools expose zero-day vulnerabilities, which are undiscovered by the software manufacturer, and APTs, which use stealth and persistence methods [47]. AI/ML systems are also being applied to identify and prevent botnets in IoT networks, as these devices' volume, variety, and openness make conventional security strategies unworkable [48,49]. Primarily, these zero-day vulnerabilities and APTs are contained in one of the quintessential aspects of AI today, the ML algorithms. The ML schemes contrast with the traditional concept of systems based on a set of rules and signatures to recognize vulnerabilities. Instead, they train on a large amount of data and build predictions to identify new never-before-seen data points as either a 'good' or 'bad' element. The capability of such recognition systems makes them especially valuable in discovering new or incipient threats that still need to be met or described [14].

However, since many of the algorithms functioning under the category of ML are 'black-box,' they may not be ideal for use in security-sensitive environments. That is why the field of XAI emerged—to explain how these models make decisions, allowing viewers to verify the conclusions independently [34,50]. Some specific examples of XAI techniques and their applications in cyber security include Local Interpretable Model-agnostic Explanations (LIME). This popular XAI technique provides local explanations for individual predictions made by any classifier by approximating the model locally with a simpler, interpretable model. In cyber security, LIME can be used to explain why an intrusion detection system flagged a particular network traffic pattern as suspicious. By highlighting the specific features that contributed to the model's decision, LIME can help security analysts validate the alert and make informed decisions about further investigation or response [50]. The application of LIME in cyber security has been explored in various studies, demonstrating its effectiveness in providing interpretable explanations for complex AI/ML models [34]. In addition, SHapley Additive exPlanations (SHAP) is another powerful XAI technique that provides local and global explanations for ML models. SHAP values quantify the contribution of each feature to a model's prediction, allowing for a deeper understanding of the factors that influence the model's decision-making process. In cyber security, SHAP can be used to explain the features that a malware classification model uses to identify a file as malicious, helping security analysts understand the underlying characteristics of the malware and develop effective countermeasures [51]. The use of SHAP in cyber security has been gaining traction, with several studies demonstrating its ability to provide meaningful explanations for complex AI/ML models [52].

Lastly, programs developed using ML algorithms keep learning from the data and are updated regularly since cyber security threats are constantly advancing. This continuing process is vital for defining and eradicating risks before the occurrence of irreversible damage [8]. One of the factors that make AI and ML favorable in the enhancement of security is their capacity to adapt and learn autonomously without supervision from people all the time. However, a wealth of evidence demonstrates AI's capacity to improve cyber security, which is supported by [53], who identify the capability of AI in redesigning threat detection for the next generation. They allow a high-security alert level to be maintained, which is far superior and less labor intensive than other methods used in the current information security technical environment. Therefore, first, it helps handle extensive data analysis successfully, and second, it makes human analysts shift their attention to activities concerning handling incidents and hunting threats.

2.3.2 Speed of Response

Utilizing AI technologies is set to drastically change threat response in cyber security. It will allow for greater automation of the threat response process, shortening the time from threat identification to the countermeasures applied. Their rapid response capacity is essential in countering the harm that can be caused by a cyber attack, especially when the time is minimal [9].

For example, imagine a case in which an AI system in an organization identifies an anomaly in the traffic in the network, probably a sign of a cyber attack. Rather than waiting for the information technology (IT) staff to respond, often with a certain amount of delay because of incident logs, the AI system can implement preprogrammed activities, including blocking probable malicious traffic, containing compromised network areas, or triggering a more extensive response procedure [37]. AI tools perform the above almost immediately after the first detection of the attacker, which severely restricts the attack window and achievable damage.

Furthermore, it relieves the pressure and workload of the cyber security teams that deal with hundreds, if not thousands, of alerts and incidents daily. Some of the tasks that may have been carried out manually, such as the response management process, escalated indicators, and executed actions, can be effectively handled by AI systems, freeing up the energy and competency of the teams to work in areas like threat hunting, vulnerability management, and security policy. This approach enhances the functionality and productivity of the cyber security process since the organizations in question have a limited number of cyber security tools at their disposal [8].

Response systems can be embedded with self-learning capabilities to prevent future occurrences by enhancing the system's ability to diagnose and fight different threats. Closely related to the first advantage, this adaptive capability is helpful in the constantly changing environment of cyber threats characterized by new attacks and methodologies. Using AI and ML in organizations can lead to the development of more robust security features, especially when protecting an organization from cyber threats.

2.3.3 Predictive Capabilities

ML models take data and discover patterns; it has become one of the most effective tools for estimating possible weaknesses and probable threats in cyber security. This proactive approach stands out from the usual methodologies that adopt a reactive stance and enables organizations to enhance their protection in risky areas before an attack [13]. Besides their reactive perspective, the models of AI and ML are primarily effective in terms of prediction. These models look at previous occurrences and recognize them as patterns, allowing for an estimated understanding of how and where an opponent could come in [54].

For example, ML models can scan through a large amount of data, including historical attacks, vulnerable information, and threat intelligence data, find the relation and correlation between these data, and recognize a system's weaknesses, such as out-of-date software, misconfigure of setting, and abnormal user actions, which attackers might take advantage of [14].

Also, by using threat intelligence data, ML enables an organization to predict which attacks are potentially imminent by analyzing the trends of malware families, phishing campaigns or any other vectors and estimate the potential impact of a threat [39].

Thus, organizations can prevent the mentioned risks while executing the change management process (e.g., if the predictive model has singled out a carrier as a likely avenue for ransomware attacks). Otherwise, the organization can raise the level of protection around that vector in advance, for example, by strengthening access regulations or expanding observational measures. They can also inform the employees of the possibility of such an attack and the measures they should take not to become a victim [9].

Such a proactive approach helps prevent potential security breaches and allows organizations to allocate resources more effectively. By focusing on high-risk areas identified by predictive models, organizations can optimize their security investments and achieve higher protection with limited resources.

2.4 Implementation Strategy

2.4.1 Assessment of Current Infrastructure

The current state of the infrastructure requires a thorough evaluation before incorporating AI in the fight against cyber threats. Thus, this assessment is an essential basis for evaluating the positive and negative aspects of the current cyber security model and determining the AI technologies that can be implemented most efficiently [9].

Thus, the assessment should effectively review the state of the organization's security in terms of the extent and kind of security control, policies, and procedures, which include security measures at the network layer of the architecture [12], at the endpoint, in handling the data, the response to any incidents, and any training on security awareness. Also, the assessment should define the characteristics of the organization's data, as this data will be instrumental in training/validating AI models [14].

Consequently, it becomes possible to understand which parts of the existing structure AI can make most effective–considering both key advantages and significant drawbacks simultaneously. For instance, if the assessment determines that the organization is weak in detecting advanced threats, AI-based anomaly detection systems can be critical [13]. On the other hand, if incident response times are an issue, AI automation would speed up response activities [46].

Also, the evaluation enables anyone to determine the extent of customization needed for the intelligent tools to fit into the existing infrastructure. Organizations should only introduce AI solutions as a general package that will solve some of the firm's problems since these applications differ in functionality. With the help of an assessment, it is possible to determine the main unsuitable integration issues, for instance, data reconciliations, limitations of legacy systems, or regulations, and consider these issues while designing AI solutions [9]. To counter a never-ending stream of threats, [55] observes in a recent review that AI needs to learn on the fly to be optimal.

Therefore, a comprehensive assessment of the current infrastructure is indispensable before embarking on AI integration endeavors. It provides a roadmap to identify areas where AI can provide the most significant value and ensure a smooth and successful integration process with other legacy systems.

Nonetheless, besides a comprehensive infrastructure assessment, a successful roadmap requires the following data science step, which precedes training and validation of AI/ML models, i.e., preparing and managing its data. It includes data acquisition, cleaning, preprocessing, and labeling. It is also essential to ensure that the data collected and used in the model is good, variable, and comes from the real-world security situation. Versioning and managing data should be done to track changes made, guarantee replication, and support cooperation [56]. After data preparation, one can choose the appropriate cyber security-related AI/ML algorithms and models for a specific task. Hence, this decision entails comparing the type of data to be analyzed, the kind of inference desired, and the computational power available. The selected models should be trained and verified based on the prepared data set, and the model's performance should be analyzed using the proper metrics [43].

AI/ML models trained per the earlier steps should be deployed and incorporated into the current cyber security environment. This may involve creating APIs or dedicated hooks whereby the AI/ML models can communicate and share data with other security tools. One of the objectives that must be achieved is ensuring that convergence does not lead to new weaknesses or eliminating efficiencies [13]. Subsequently, the AI/ML models after deployment should be regularly examined to assess their efficiency in removing bias and performance downfall. Besides, this ongoing examination ensures the audience that the system constantly improves and adapts to new challenges. It is necessary to include the feedback provided by security analysts and other stakeholders to validate the models in the interest of aligning them to an organization's security objectives and goals [39]. The trends in cyber security are changing, with new threats and implementation methods appearing from time to time. AI/ML models must be updated and upgraded so the models remain relevant against the newer incoming potential threats. This constitutes the constant update of the training data, the training of the models, and the introduction of new AI/ML techniques where necessary [57].

Lastly, some of the best practices for AI/ML integration encourage the interaction of cyber security professionals with data scientists, as well as other parties involved in implementing AI/ML systems. This collaboration is beneficial and essential as it provides practical leadership for explaining the endeavors, constraints, and possible advantages of AI/ML implementation to stakeholders. Also, the incorporation of XAI methodologies to give human understanding and explain what AI/ML systems are doing to offer decisions. This will also make people begin to have confidence in these technologies and make it easier for security analysts to analyze the output from the model and ensure that it is correct [56]. Besides, incorporating human sight will boost the AI/ML decision-making process. It will assist in addressing probable biases, reviewing how decisions were reached, and then deciding on the security occurrences [58].

Consequently, infrastructure readiness is the first crucial step toward incorporating AI in cyber security. This ensures that AI is used where it can bring the most value, integrated with existing systems, and where there are issues. It is critical to explain how the data should be stored, how many models exist, and how cyber security and data science consultants may cooperate. Receiver updates and ongoing infrastructure maintenance will enhance the economic security of AI systems.

2.4.2 Development of AI Solutions

Based on the needs found during the infrastructure assessment, organizations can either build AI solutions independently or purchase them from reliable suppliers. Internal development proved advantageous since it would incorporate a specified solution that responds to the organization's security needs; however, engaging in internal development takes much time and adequate skills. Conversely, buying AI solutions from vendors is cheaper and faster since the solutions are pre-trained to work with the configurations acquired from other vendors. However, IT experts must tune the vendor solutions to fit the organization's infrastructure since they must integrate with the existing systems [9].

In any case, the development and integration phases focus on configuring the AI to address the identified security requirements, ensuring it fits the organization's security goals. The results must offer congruent solutions, integrate AI with other related applications, and follow different security frameworks and policies. In addition, there is a need to set up effective monitoring and evaluation regimes to track the AI's functioning, check for signs of bias or failure, and increase the AI solutions' efficiency in the long run [13].

2.4.3 Integration with Existing Systems

Incorporating AI solutions with other resources required in cyber security means making proper arrangements to create synergy. The process may include improving current systems or software to meet the computational and data processing abilities displayed by an AI algorithm, as Jordan et al. [59] mentioned. Where there are mainframe applications, compatibility, and integration, such as when passing data from one application to another, may require the development of unique bridges or wrappers [57].

Interfaces and protocols must conform with other cyber security instruments and practices to allow for maximum compatibility and enable the cohesiveness of AI with other elements. The efficient definitions of data formats and channels guarantee data input to AI systems and the availability of the results obtained for the other security tools that need them to work efficiently and without delays [60]. Moreover, there is a need to define tasks for AI and human personnel when organizing cyber security-related activities. It is appropriate to consider AI as an instrument that enriches human decision-making by offering valuable conclusions and performing routine complex calculations. Thus, human assets analyze the outcomes of the corresponding AI algorithms and make critical security decisions [58].

2.4.4 Training and Adaptation

Staff training is critical in achieving the best results from applying AI systems in cyber security. Training should include knowledge of the new tools' usage and added understanding of what AI produces and why [5]. Technical staff require training to assess AI-generated suggestions, distinguish between true alarms and noise, and act according to the organization's purposes and security goals [58]. This learning process entails changing the staff's perception from a technical one to an organizational cultural orientation toward incorporating AI in the implementation of cyber security.

Moreover, the connection of AI to operation is not a singular action but a continuous process. ML-based AI systems especially need constant updating and adjustments because the system acquires data from real-world applications and can adapt to new threat profiles, new methods of attack, and various changes in the organization's IT environment [57]. Some requirements for maintaining AI's efficiency are constantly updating training data, models, and decision thresholds due to the changes in threat scenarios. Also, human analysts should integrate their feedback into the AI systems to improve the AI's decision-making process and ensure it is in tandem with human professionals.

2.5 Case Studies

2.5.1 Successful Implementations

Integrating AI within cyber security is possible and paves the way to conquering contemporary threats. Some successful cases include a flagship financial institution that applied and adopted AI in the cybersphere by developing ML algorithms, which increased fraud detection and decreased the rates of false positives to a large extent [61]. Extensive transactional data passed through the trained ML models, and complex patterns and potential indicators pointing to fraud appeared, meaning that the AI tools worked as expected. Therefore, these results show that even with more advanced fraud schemes, AI can track them straightforwardly, but rule-based systems fail to do so.

In addition, identifying fraudulent transactions through the AI-based system ensured that applicants' transactions were flagged appropriately and reduced false positives that would be detrimental to the institution's security [62]. It also offered better customer situations by lowering the inconvenience and interruptions caused by inaccurate security alarms. Thus, this case reflects that AI creates opportunities for financial cyber security, improving the industry's security level, increasing operational productivity, and facilitating customer satisfaction.

The Canadian Institute for Cybersecurity (CIC) created an Artificial Intelligence Network Intrusion Detection System (AI-NIDS) to detect cyber threats such as DoS, DDoS, Brute Force, XSS, SQL injection, and infiltration attacks [63]. Using deep learning models, the NIDS also demonstrated a high level of accuracy in detecting unwanted network traffic, even if the patterns of the attacks are novel or previously unseen. The NIDS also demonstrates how AI can improve network security by identifying multiple types of cyber attacks with high performance and flexibility.

Furthermore, Saxe et al. [64] concluded that researchers at MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) created an improved form of the AI2 system, dramatically raising the level of malware detection. AI2 trained itself to detect malicious code. Hence, it depended entirely on the behavior and structure of the fed code rather than depending on the virus's signature. Thus, this approach enabled AI2 to detect novel malware variants and zero-day attacks with greater accuracy and speed than traditional antivirus software.

In addition, in the healthcare industry, AI is being used to improve cyber security in the following manners. A notable example of this is the ML-based monitoring of the behavior of medical devices to identify signs of cyber attack or mechanical failure [65]. It is preventive and assists in maintaining safety for patients while enhancing the quality of sensitive healthcare systems. Similarly, researchers have successfully applied data augmentation to improve the effectiveness of AI/ML systems in the classification of tumors [66], and it can be used similarly in malware detection [64].

AI and ML are also increasingly used to protect critical infrastructure, including power plants and transportation. For instance, ML algorithms are being used in the power grid to monitor sensor data to check where intrusion detection is or some form of irregularity, suggesting that a cyber attack is imminent [67]. Hence, this technique allows for early identification of a threat and its containment so that any effects on main functions can be limited.

Lastly, the retail industry implements AI to fight e-commerce fraud and safeguard customers' information. ML algorithms forecast customers' behaviors and variance in transactional trends so that they can be assessed for risk factors characteristic of fraudulent transactions [68]. AI assists in minimizing financial risks so that customers' confidence and loyalty will not be misplaced.

The above case studies and the existing real applications demonstrate the diverse uses of AI and ML in cybersecurity across various industries. However, continued success relies on addressing challenges such as the need for XAI, as emphasized by Ribeiro et al. [50], to ensure transparency and trust in AI-driven decision-making. The following section explores these and other lessons learned from real-world AI implementations.

2.5.2 Lessons Learned

The following challenges persistently emerge whenever AI solutions are deployed in cyber security contexts, complicating these technologies' integration and use. Privacy, disparate data quality problems, qualification, oddity, insufficiency, or inherent biases can affect AI models, among other things. For example, if the training data set differs from real-life cases, the AI will need help generalizing the learned data to serve real-life applications [69]. Managing these data quality issues requires data cleansing, where errors, missing values, and irrelevant data are handled and solved; transformation, where all the data is normalized and feature scaled; and data augmentation, to diversify the training dataset.

Interoperability remains a significant issue because of the heterogeneity of cyber security systems applied in the existing structure, which comprises disparate tools and infrastructures. Sometimes, these comprise legacy systems and other proprietary software applications that may require the production of interfaces, wrappers or brokers to link them to AI solutions [13]. Adequate planning, technical experience, and active collaboration from IT and cyber security personnel can solve the above implementation issues and guarantee a smooth and efficient security ecosystem operation.

Moreover, a significant challenge is a reluctance to change to new technologies and products, especially among employees, who must be more proactive regarding the influence of artificial intelligence management (AIM) on their work. This organizational resistance can lead to organizations' unwillingness to adopt new AI tools and platforms, doubts about the efficiency of AI-based recommendations, or even flat-out rejection of AI-based processes [70]. Consequently, planning, effective communication, and technical expertise must work correctly for change management. Organizations should design communication strategies about AI's pros and cons, provide opportunities for enhancing employees' knowledge about AI tools, and increase their awareness and trust about the benefits these technologies offer.

2.6 Challenges and Ethical Considerations

The increasing use of AI in cyber security also raises important ethical considerations. It is essential to carefully address issues such as the potential for bias in AI algorithms, the risk of discrimination, and the misuse of AI-powered tools for malicious purposes [71]. The challenges of integrating AI into cyber security are technical and involve organizational and cultural factors. References [72,73] provided a comprehensive overview of these challenges, including theoretical approaches and practical solutions for mitigating them. Additionally, the potential negative impacts of AI in cyber security, including the risk of adversarial attacks and the need for robust security measures to protect AI systems themselves, have been highlighted by various studies [73,74].

There are several factors that the literature has singled out as barriers to the implementation and usage of AI/ML for cyber security solutions [72,73]. The performance of AI/ML models significantly depends on the availability and quality of data used to train the models. Collecting massive amounts of good quality, labeled data in cyber security is challenging because security information is often sensitive, while threats are ever-dynamic. A lack of labeled data can also cause models to undershoot or overshoot the optimal level of generalization, affecting the performance of the models. Moreover, an uneven distribution with rare classes, such as malicious activities compared to other classes, will be prevalent, making the training mechanism skewed. To deal with the aforementioned data-related issue, according to the literature, there should be substantial data collection and labeling processes [5,9,14]. Current security systems and structures are heterogeneous and complex, and it is challenging to integrate AI/ML solutions to overcome security loopholes. AI/ML models' integration with existing structures, security systems, and processes entails a comprehensive strategy, a profound understanding of IT, and cooperation between different departments. Some security solutions may not be compatible with others, which can cause more problems as they attempt to integrate them. Regarding the requirements, the literature emphasizes developing convenient and scalable AI/ML solutions that integrate into current security systems [13,57].

Most AI/ML models, especially those using deep neural networks, are called 'black boxes' because their inner workings are complex. In cyber security, others must be confident in the model and its recommendations. Thus, the lack of interpretability of the model's decision process may prevent the usage of the model. This problem is still a hot research issue in the context of the growth of a new scientific branch called XAI, which focuses on creating methods that put the user in front of the model's decision-making process [34,50,56]. However, more investigations are warranted to establish proper XAI methodologies that best work in the cyber defense landscape.

Furthermore, AI/ML also brings fundamental shifts in how cyber security is attempted and defended and how it relates to organizations and their personnel. Some challenges associated with implementing AI/ML in cyber security include a lack of expertise within the cyber security teams that form the human aspect of AI/ML resistance to change by employees and the management. The members of organizations must support change through communication and training plans. Regarding AI/ML, organizations must create a workplace with employees possessing these skills [70]. Another critical barrier is that AI/ML-based cyber security solutions, like any other technology solutions, require development, implementation, and management resources, which could include time, money, and expertise. There will be limitations to the extent that smaller organizations or organizations with limited funding may find it challenging to implement these technologies [9]. Resource scarcity means supporting low-cost strategies for outsourcing AI/ML services from the cloud or freeware [9]. Moreover, AI and ML can be misused to create highly advanced cyber threats or botnets capable of spreading fake news. Besides, this reality means that there should be set principles and policies that check on the use of AI/ML to avoid being used for other purposes but should be used to improve the welfare of society [58].

Solving these technical and organizational concerns requires multiple approaches involving new technologies coupled with organizational changes. When such issues are handled in advance, the benefits of applying AI and ML to boost an organizational cyber security strategy can be realized. In addition, Reference [75] drew focus on ethical issues regarding the application of AI in cyber security and how to reap the AI benefits while abiding by ethical standards. Applying AI in cyber security has its drawbacks despite its functioning. One issue is that many of the AI models' underlying algorithms seem to be a 'black box,' meaning it is challenging to understand the decision-making process of AI models [56]. The lack of explainability of the model's thinking process presented in this paper can hinder trust in AI-based security tools. It may make it challenging for human analysts to understand why the model took a particular action.

Applying AI to cyber security implies analyzing large volumes of distressed information, which leads to privacy issues. Sometimes, such data involves Personally Identifiable Information (PII), Intellectual Property (IP) or other corporate-sensitive data. Errors or breaches of such data can lead to money loss, reputation impairment, and legal consequences [76].

AI integration is also effectively applied in cyber security; however, this development has several ethical concerns. There are questions regarding bias in AI systems, the problem of discrimination, and the threats related to using AI in criminal activity or other wrongdoing [71]. Notably, integrating AI into cyber security is not purely a technological problem but has organizational and cultural aspects. These challenges are showcased in detail in [72,73,77,78] outlining theoretical frameworks and essentially practical ways to address them. Also, some authors have discussed the adverse effects of AI in cyber security in several works regarding adversarial attacks, which make security measures necessary to protect AI systems [73,74].

Consequently, it is urgent to devise rules of ethics and standards for using AI in cyber security to avoid adverse effects. These actions range from encrypting the data at rest and in transit, anonymizing or pseudonymizing the data, and regulating data access [79]. It also has a legal and regulatory obligation aspect, such as the GDPR standards that govern data processing and uphold privacy [1]. In this regard, organizations must adhere to privacy laws like the General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA). Moreover, data anonymity and de-identification methodologies should be strengthened within organizations to safeguard a person's identification [39].

Bias in training data forms another challenge that hinders both the efficiency and fairness of the solutions we have for AI in cyber security. Thus, if the training data has biases, so will the outcome of further work on the AI model, and discriminatory results or low prediction accuracy will emerge [54]. The biased AI model might tend to report the actions of people belonging to particular groups as suspicious and, hence, victimize the said groups. Organizations must employ diverse and balanced datasets to train the AI systems to address this risk and avoid skewed information towards a particular third party. Also, fairness-aware methods and annual bias checks and calibrations are recommended [54]. Such models require constant updates to prevent biases and conflict with the principles of equality and fairness in using AI in cyber security. As with all models that use AI and ML techniques, the quality of the data used to train them is crucial if they are effective in cyber security [9,14]. Bad data quality will cause wrong predictions, false positives, and false negatives, ultimately resulting in a weak security instance. The importance of data quality for cyber security applications is evident, as seen in several studies, both in terms of the effect on the performance and the trustworthiness of AI/ML models [1,5,9,14,26]. A recommended way to enhance data quality is to deeply clean and preprocess the data. This action includes dealing with missing values, outliers, noisy data, data consistency, and data integrity. Imputation, outlier detection, and data normalization are methods that deal with these problems [69]. Sarker et al. [9] have stressed the role of data preprocessing for cyber security applications, which shows how it can enhance the accuracy and performance of AI/ML models.

Lastly, AI systems and technologies are also at risk of cyber threats. Potential attacks can be made on AI algorithms or training data, making AI-based security systems vulnerable [80,81–85]. AI needs protection like any other complicated computer-based system, including firewalls, intrusion detection systems, and access controls to avoid unauthorized access to AI systems. Continuous training of AI models and software must incorporate fixes for the vulnerabilities in the market. Some of the best practices and guidelines that organizations might implement to avoid ethical and legal issues [86] include systems that are easily understandable and capable of providing precise and accountable details; accurate AI/ML systems that mitigate prejudice risks; and effective and appropriate measures in AI and ML decision-making to minimize negative impacts. According to [86], sound management also implies that if there is continuous monitoring of the functioning of AI, one would be able to identify any irregularities or any activities deemed suspicious, which would mean there is a possibility of breaches in the AI systems.

2.7 *Cyber Security and Emerging Technologies*

With new advancements in AI and ML at a very high frequency, further changes are happening in this respective cyber security segment. Since organizations must face more complicated threats, incorporating AI and ML can become one of the ways of enhancing security. Recent developments suggest that the role of AI-based solutions is crucial in addressing, identifying, and mitigating threats [71,87]. Future developments of advanced varieties of computing, such as quantum computing or higher levels of NLP, only serve to increase this optimism [28,69,88].

Nevertheless, achieving total value from AI and ML in cyber security depends on several key factors: the quality of input data, the integration issue, and some ethical aspects discussed in [56,89]. However, further development is required to investigate the potential of new applications of these technologies in the context of the latest threats and risks. This paper has drawn from [90] and [86], advancing the idea that constant research and development of AI and ML are essential for the future of cyber security. Furthermore, References [2,3] presented the existing threats in the field of cyber security, and Reference [40] focused on the IoT world's opportunities and threats. All these challenges point to the fact that there is a need to continue with the development of new AI solutions to meet the dynamics of the current threats.

3 Methodology

This systematic literature review analyzed 90 studies on AI and ML applications in cyber security following an information systems research protocol by Okoli et al. [91], as in Fig. 1, and using databases such as IEEE Xplore, ACM Digital Library, ScienceDirect, Web of Science, and Scopus. These databases were chosen for their comprehensive coverage of peer-reviewed articles, conference proceedings, and computer science and engineering technical reports.

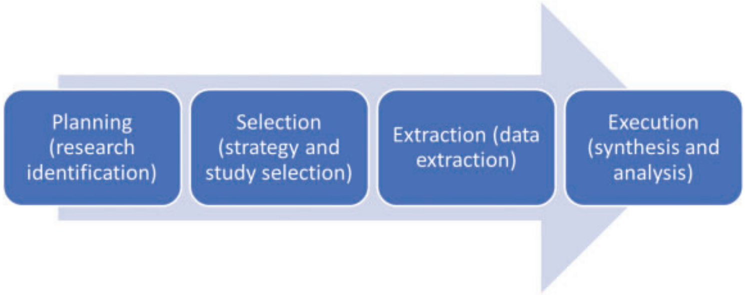
Planning (research identification)
Selection (strategy and study selection)
Extraction (data extraction)
Execution (synthesis and analysis)

Figure 1: Steps followed in the literature review. Note: Taken from [92]

Furthermore, Table 1 shows the critical elements of Okoli et al.'s [91] protocol.

Table 1: Okoli and Schabram's key elements

Element	Concept
Formulating the research question	Clearly defining the research question or objective of the SLR. This will guide the entire review process and determine the scope of the literature search.
Defining inclusion and exclusion criteria	Establishing specific criteria for selecting relevant studies. This includes factors such as publication dates, types of studies, and research methodologies.
Conducting a comprehensive literature search	Systematically searching relevant databases, academic journals, and other sources to identify all potentially relevant studies.
Screening and selecting studies	Applying the inclusion and exclusion criteria to the identified studies, carefully screening them for relevance and quality.
Extracting and analyzing data	Extracting relevant information from the selected studies and analyzing it systematically. This can involve qualitative or quantitative methods or a combination of both.
Synthesizing the findings	Integrating the findings of the individual studies to answer the research question. This involves identifying patterns, themes, and inconsistencies across the literature.
Reporting the results	Presenting the findings of the SLR clearly and concisely. This includes describing the methodology, summarizing the findings, and discussing future research and practice implications.

In the planning phase, the main research question is: How can AI and ML be effectively integrated into cyber security practices to enhance threat detection and response mechanisms while addressing associated challenges and limitations? In addition, regarding the main research question, several questions required further analysis:

1. What challenges and limitations are associated with integrating AI and ML into cyber security, and how can they be addressed?
2. How can XAI ensure transparency and trust in AI-driven cyber security measures?
3. How can organizations overcome data quality and integration complexities when implementing AI and ML solutions in their cyber security frameworks?

In the selection phase, a comprehensive qualitative search strategy ensured the inclusion of relevant studies. This strategy combined keywords and Boolean operators (AND/OR) to capture a wide range of research on AI and ML in cyber security. The primary keywords included "artificial intelligence," "machine learning," "cybersecurity," "threat detection," "network security," "vulnerability assessment," "intrusion detection," "malware analysis," "phishing detection," and "security automation." The search covered the materials comprehensively using IEEE Xplore, ACM Digital Library, Science Direct Web of Science, and Scopus. In the extraction phase, specific inclusion and exclusion criteria maintained the review's focus and relevance. The review included studies that met the following criteria: (a) published in English between 2015 and 2024, (b) peer-reviewed articles, conference papers or technical reports, (c) focused on the application of AI or ML techniques in cyber security, and (d) provided empirical evidence or theoretical frameworks supporting their findings. Studies not published in English, focused on AI/ML applications outside of cyber security, or lacking empirical evidence or theoretical grounding were excluded. The period from 2015 onwards marks a significant acceleration in the development and application of AI and ML technologies. Fundamental breakthroughs in deep learning and other AI techniques occurred during this time, making it a crucial period for studying their impact on various fields, including cyber security. Cyber threats have evolved rapidly over the past decade, becoming more sophisticated and widespread. Focusing on this timeframe allows the study to capture the latest trends and advancements in cyber threats and the defensive measures developed to counteract them.

The identified studies were reviewed according to set standards to determine their quality and suitability for inclusion in this review. The appraisal process involved two main stages:

1. Initial Screening: The inclusion and exclusion criteria were used to scrutinize only the titles and abstracts of the identified studies. At this stage, any paper that did not describe the use of AI or ML for security or was not backed up by empirical studies or theories was eliminated.
2. Full-Text Review: The full texts of the remaining studies were examined to evaluate their methodological quality. Issues covered included the kind of study, sample size, method of data collection, data analysis, and source of bias. Cohort studies with low methodological quality or that met specific exclusion criteria were finally omitted. The 90 articles finally included represent a comprehensive and diverse set of research articles covering AI and ML use in cyber security, ensuring a broad perspective in the review. The review extracted critical data from each identified study, including study objectives, methodology, results, and conclusions. This data was organized into a standardized format to facilitate analysis and synthesis. A critical appraisal then assessed the included studies' methodological rigor and overall quality, considering factors such as study design, sample size, data collection methods, analysis techniques, and potential biases. The quality assessment ensured the literature review findings relied on credible and reliable evidence.

The extracted data revealed prevalent themes in the literature, which will be further analyzed and discussed in the results section of this research. Fig. 1 illustrates the distribution of publications across the years, highlighting a significant increase in studies on AI and ML applications in cyber security.

The bar graph in Fig. 2 shows a clear upward trend in publications, indicating a growing interest in and research into AI/ML for cyber security. The data suggests that the field of study was relatively new in the early years (2015–2017) and then experienced a surge of interest and development. The decline in 2024 might indicate that the field is maturing, with research focusing on more specific and specialized areas rather than broad exploratory studies, as seen in the next section. Also, it is worth noting that several papers might still be in the publication process, and end-of-the-year statistics could show a different trend. Nonetheless, it is worth noting that almost ten years of data provide a good overview. Nevertheless, it is a relatively short period. Longer-term trends might reveal a different picture. In addition, several references such as [9,25,27,69] were relevant to multiple themes in this analysis, highlighting the interconnected nature of AI/ML and cyber security research.

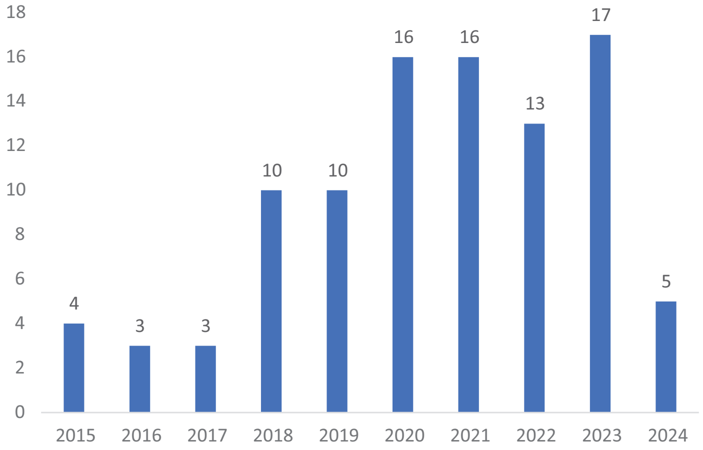

Figure 2: Total number of publications by year

4 Results and Discussion

In this analysis of the AI/ML for cyber security research landscape, the publications were categorized into seven distinct themes: Data Analysis and Automation, Fundamentals of AI and ML in Cyber Security, Network and System Security, Challenges and Limitations, Future Directions and Emerging Technologies, Current Cyber Security Challenges, and Threat Detection and Prevention. These themes were selected based on prevalent topics identified in the literature review, reflecting key areas of focus and emerging trends in the field. Fig. 3 summarizes the total number of publications for each theme.

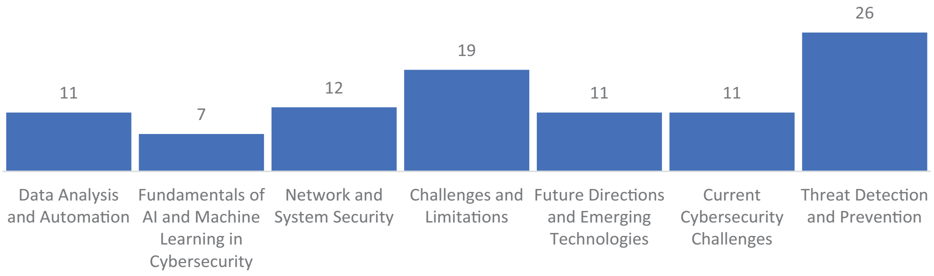

Figure 3: Total number of publications by themes

As the bar graph in Fig. 3 shows, the distribution of publications in AI/ML for cyber security research reveals a diverse landscape with varying emphasis on different themes. Threat detection and prevention are the most prominent themes, comprising 27.2% of the publications. Threat detection and prevention are followed by challenges and limitations (19.6%), future directions and emerging technologies, data analysis and automation, and current cyber security challenges, each accounting for 12%. In contrast, the theme of fundamentals of AI and ML in cyber security and networks and systems security receives less attention, with only 6.5% and 10.9% of publications, respectively. Fig. 4 shows the percentage distribution of publications by theme.

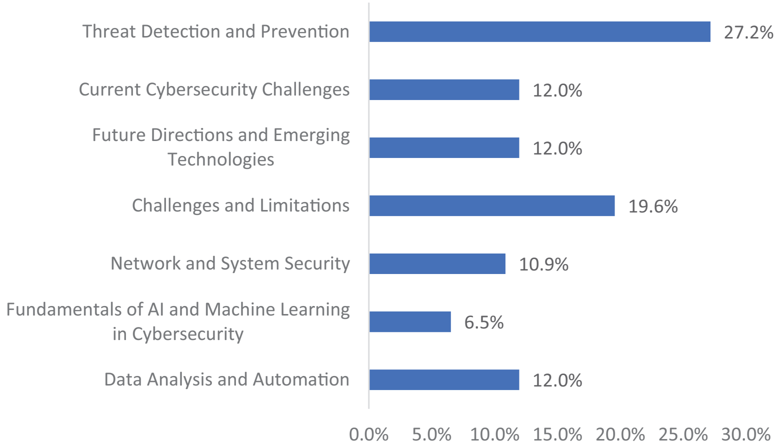

Figure 4: Distribution of publications by themes

These trends suggest that while the field is actively seeking ways to apply AI/ML to real-world security problems and explore emerging technologies, there might be a need for more foundational research and a greater emphasis on data-driven automation and networks and systems security to realize the potential of AI/ML in cyber security fully. The relatively few publications on foundational aspects and data automation could suggest a gap in the research landscape. The relatively few publications on foundational aspects and data automation suggest a gap in the research landscape. While there is a strong focus on applying AI/ML to specific cyber security problems, there might be less emphasis on the theoretical underpinnings and the automation of data-driven processes, which could hinder the development of robust, reliable, and explainable AI/ML models.

The literature review reveals a burgeoning interest in applying AI and ML to bolster cyber security defenses. A significant portion of research focuses on threat detection and prevention, with studies exploring diverse AI/ML techniques to identify and mitigate various threats, including malware, phishing, and zero-day attacks [5,8,9,13,14,27,34,38,43,56,61,70,90,93]. This emphasis underscores the pressing need to address the evolving threat landscape and the potential of AI/ML to enhance existing security measures.

Network and system security is another prominent area of research, with studies investigating the use of AI/ML to secure networks, endpoints, cloud environments, and IoT [13,39,40,46,87]. These studies highlight the importance of adaptive and intelligent security measures to protect critical infrastructure and interconnected devices from increasingly sophisticated attacks.

However, this paper's literature review also denotes the risks and limitations of adopting AI/ML in cyber security. Concerns from the literature include data quality issues, integration challenges with other systems, and XAI [2,3]. However, AI models can be biased, and adversarial attacks are still among the issues that need further investigation and improvement [54,80,94]. Nonetheless, there are some challenges when it comes to integrating AI and ML that work for cyber security, and in this case, there is a promising future. Various works by authors like [93] look into the future, possibilities, and usage of AI and ML-based cyber security, contemplating the concerns and issues of the future.

Moreover, the literature reveals that features such as quantum computing and other advanced techniques such as NLP can reinforce the outcomes of AI-based security measures much further [69,88]. Also, studies on Federated Learning (FL) [89] and AI for security automation [9] present new opportunities to solve privacy issues and the scarcity of cyber security talent.

The literature and journals presented suggest that the scope of AI/ML is broadly interdisciplinary, as the journal titles of the studies also indicate. The variety presented underlines the need to discuss such a field with multiple scholars—not only computer scientists and engineers but also scholars of the social sciences.

The literature review generally reveals a rather diverse and growing field, focusing on practical application and technological advancement. Although scholars have attained a considerable level of advancement, more research efforts shall be employed to address the mentioned challenges and to unfold other types of research that will enhance the optimization of AI and ML in cyber security.

5 Future Research Directions

One of the most significant issues is that the most popular AI models are 'black boxes,' and the reasons behind some of the decisions may not be obvious. Therefore, an immediate research priority should be XAI models able to explain their actions, which help to increase the trust in AI [56], ensuring the creation of new AI systems that are immune to adversarial attacks; Further, for the future research, it is essential to analyze the application possibilities of the novel technologies and the development of the sophisticated NPLs, as well as to consider the ethical and the policy issues corresponding to the use of the AI in the sphere of cyber security [85,90]. Moreover, it is critical to invest in research to strengthen the AI models used in defending security applications against adversarial attacks like data manipulation and model obfuscation [28].

The growing awareness of data control and protection requires using distributed and private machine-learning systems that can learn from distributed data. One of the promising techniques for collaborative learning across multiple devices or organizations that maintain localized data is FL. Therefore, the next research tier should investigate FL's capability in threat detection, malware analysis, and anomaly detection, as Yang et al. [89] discussed in their work. This fact, as well as the constantly growing evolution of cyber threats and the lack of workers with security expertise, contribute to why there should be more automation in security operations. Subsequent studies should integrate AI in vulnerability assessment, patching, and incident handling so that researchers can focus on higher-level work [9]. This study aligns with the research pointers espoused by [95] on the possibility of AI/ML incorporation under security orchestration automation and response.

Quantum computing is viewed as a game changer in the field of AI concerning cyber security, since new advanced algorithms enhance the identification and analysis of risks. The study of quantum ML and its capabilities in cyber security content is an emerging yet vital research direction [88]. Also, the advent of 5G networks offers new AI security and threat prevention opportunities. Research should urgently address issues related to more significant exposure to attack, threat detection in real-time, and network slicing security [87,81–85].

Based on multiple studies explored in this paper, the use of AI and ML in cyber security has numerous improvements to threat detection, response, and prevention. However, this literature review also has some weaknesses that should not go unnoticed.

Firstly, due to the nature of the sources, the articles included only those in English, so there may be significant studies in other languages or other countries outside this review. This limitation may impact the comprehensiveness of the review. Thirdly, there is a possibility of being restricted to peer-reviewed journal articles only and, therefore, may fail to account for the brilliant revelations that could be provided by the conference papers, technical reports, and industry white papers, among others.

It is also noteworthy that the quality of the included studies is debatable: some of the articles might have methodological flaws or contain biases that might influence the results. Next, because outcomes can vary, this also must be considered when using them. Moreover, the given results are limited in their generalizations because some AI and ML applications may be related to specific industries or types of cyber threats.

Given the rapidly evolving nature of AI and ML technologies, some findings need to be updated quickly. The literature review may need to fully capture emerging threats and recent technological advancements. Moreover, while the review discusses the technical aspects of AI and ML integration, it may not comprehensively address the ethical and legal considerations crucial for successful implementation.

Notably, the derived and used AI is predicated on potent, large, and diversified data, which may only sometimes be available. Lastly, most reviewed studies formulate a significant need for integrating AI and ML in cyber security; however, the integration is a complex process that demands a combined technical, organizational, and ethical approach not discussed in detail in the reviewed literature.

Even with such limitations, the review's central message about the future of cyber security with AI and ML is still valid. Thus, future research should work to eliminate these limitations through more diverse studies, interdisciplinary approaches, and frequent data updates to include new technologies and threats.

6 Conclusion

Based on this overall literature review, AI and ML show the possibility of advancements in anomaly detection, malware classification, and preventing phishing attacks. Whereas rule-based methods are more effective in countering simplistic malware and viruses, AI and ML have evolved as the most potent weapons in today's cyber security solutions, given their ability to process vast amounts of data, find patterns, and consider the dynamism of threats.

Real-world examples discussed in this paper include examples of the use of AI for threat identification in real-time, as well as cases of the implementation of comprehensive incident response using AI, which proves the effectiveness of the solutions and the improvement of security in organizations. Thus, these case studies also present problems, including data quality, integration issues, and possible biases in AI models that require solutions for effective implementation.

Thus, the author underlines the need for an extensive approach to implementing AI, specifically ML, in cyber security. Technological solutions are insufficient; one must consider change and organizational and ethical implications. Overcoming these challenges includes enhancing compliance with legislation and regulation requirements and cooperating between human analysts and AI systems. If these complex issues are solved, organizations can benefit from the advantages AI and ML offer against cyber threats to the optimum level.

In addition, Al and ML sources also show that they provide not only a response-based, but also a proactive approach, displaying possible weaknesses or probable threats. Such prediction is essential as it arms an organization with the means to strengthen its protection ahead of time and to spend the budget to achieve a higher level of protection.

In the future of cyber security, maintaining organizations' security will involve synergy between human talent and AI technology. Altogether, it is possible to underscore that organizations that follow this integration and solve the encountered issues can achieve a much stronger cyber security level. Thus, AI and ML will continue to evolve with relevant threats and their responses and prevention, placing them at the epicenter of the fight against cybercriminals.

Therefore, AI and ML aid in improving cyber security; however, this enhancement should rely on systematically implementing innovative technologies involving technical, organizational, and ethical aspects. Organizations that consciously adopt these technologies and work on mitigating these issues will be in a much better position to deal with future cyberspace threats. The accomplishment toward the actualization of AI/ML in the cyber security domain is still in its progressive phase. Still, such strengths articulated in this review constitute solid arguments for the continuous development and implementation of AI/ML in cyber security.

Researchers should endeavor to develop unprecedented ways of expounding AI/ML decisions, enhancing confidence in AI/ML, and enhancing human-AI symbiosis. Given the ongoing attacker-defender dynamics since the beginning of this topic, researchers need to study further how to defend AI/ML models against adversarial attacks. Therefore, it is imperative to arrive at suitable methods for identifying, neutralizing, and discouraging such an attack as much as possible to advance the sensitivity of AI/ML-based security measures. As more users are aware of the risks of exposing their data and more AI/ML models must be trained on distributed data, further study of FL and other privacy-preserving AI methods in cyber security is required. Researchers should conduct more studies to determine how to learn and share threats while protecting classified information.

The recently accelerated growth of both the scale and the level of cyber threat means that security must become increasingly automated. Authors should pay more attention to ML algorithms that would support the automation of mundane tasks, contribute to the generation of early real-time alerts, and leave human expertise to decision-making tasks. Modern technology development includes 5G, quantum computing, and IoT, which push the bar regarding security threats. On the same note, researchers should perform adequate research on applying AI/ML technologies to protect these emerging technologies; researchers should develop research and innovative measures for dealing with vulnerabilities in such technologies. Any advancement of A/ML in the cyber security domain requires that there should be acknowledged and accepted ethical and policy guidelines. More importantly, researchers should develop guidelines and proper approaches to the correct and ethical use of AI/ML to promote transparency, accountability, and fairness.

Lastly, cyber security practitioners might deploy standard data cleansing, data preparation, and management practices to ensure high data quality and the resulting AI/ML models. There is a need to choose carefully the AI/ML algorithms and models to work for different cyber security tasks using the data with their specific features. Organizations should invest sufficient time in validating the models, testing their usability for the respective functions, and ensuring the use of XAI methods to enhance understanding of AI/ML in the decision-making process among security analysts and build their trust in the systems and promoting information sharing between information security departments, scientists, and other stakeholders involved in the AI/ML implementation process for better integration. Practitioners also provide cyber security workforce training and development programs to ensure the necessary AI/ML competencies for effectively managing and using these tools. Organizations might integrate moral standards and norms in implementing and designing AI/ML programs to avoid non-discriminatory, transparent, and accountable operations and to follow privacy principles.

Acknowledgement: None.

Funding Statement: The author received no specific funding for this study.

Availability of Data and Materials: The author confirms that the data supporting the findings of this study are available within the article.

Ethics Approval: Not applicable.

Conflicts of Interest: The author declares that they have no conflicts of interest to report regarding the present study.

References

[1] S. B. Atiku, A. U. Aaron, G. K. Job, F. Shittu, and I. Z. Yakubu, "Survey on the applications of artificial intelligence in cyber security," *Int. J. Scient. Technol. Res.*, vol. 9, no. 10, pp. 165–170, 2020.

[2] R. Sen, "Challenges to cybersecurity: Current state of affairs," *Commun. Assoc. Inform. Syst.*, vol. 43, no. 1, pp. 22–44, 2018. doi: 10.17705/1CAIS.04302.

[3] L. Wang and XV. Wang, "Challenges in cybersecurity," in *Cloud-Based Cyber-Physical Systems in Manufacturing*, Cham, Switzerland: Springer, 2018, pp. 63–79.

[4] M. Husák, J. Komárková, and E. Bou-Harb, "Survey of attack projection, prediction, and forecasting in cybersecurity," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 647–664, First quarter 2018.

[5] T. D. Wagner, K. Mahbub, E. Palomar, and A. E. Abdallah, "Cyber threat intelligence sharing: Survey and research directions," *Comput. Secur.*, vol. 87, no. 4, Oct. 2019, Art. no. 101589. doi: 10.1016/j.cose.2019.101589.

[6] Verizon, "Data Breach Investigations Report," 2023. Accessed: Jun. 16, 2024. [Online]. Available: https://www.verizon.com/business/resources/reports/dbir/

[7] D. Shu, N. O. Leslie, C. A. Kamhoua, and C. S. Tucker, "Generative adversarial attacks against intrusion detection systems using active learning," in *Proc. 2nd ACM Workshop Wireless Secur. Mach. Learn*, Jul. 2020, pp. 1–6.

[8] Cisco, "The role of machine learning in cybersecurity," 2023. Accessed: Jun. 16, 2024. [Online]. Available: https://www.cisco.com/c/en/us/products/security/machine-learning-security.html

[9] I. H. Sarker, A. S. M. Kayes, and P. Watters, "Cybersecurity data science: An overview from a machine learning perspective," *J. Big Data*, vol. 7, no. 1, 2020, Art. no. 41. doi: 10.1186/s40537-020-00318-5.

[10] A. J. G. De Azambuja, C. Plesker, K. Schützer, R. Anderl, B. Schleich and V. R. Almeida, "Artificial intelligence-based cyber security in the context of Industry 4.0—A survey," *Electronics*, vol. 12, no. 8, 2023, Art. no. 1920. doi: 10.3390/electronics12081920.

[11] L. Xiao, X. Wan, X. Lu, Y. Zhang, and D. Wu, "IoT security techniques based on machine learning: How do IoT devices use AI to enhance security?" *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 41–49, 2018. doi: 10.1109/MSP.2018.2825478.

[12] M. Rizvi, "Enhancing cybersecurity: The power of artificial intelligence in threat detection and prevention," *Int. J. Adv. Eng. Res. Sci.*, vol. 10, no. 5, 2023. doi: 10.22161/ijaers.105.8.

[13] H. Xu, J. Wu, Q. Pan, X. Guan, and M. Guizani, "A survey on digital twin for industrial Internet of things: Applications, technologies, and tools," *IEEE Commun. Surv. Tutor.*, vol. 25, no. 1, pp. 123–145, Mar. 2023. doi: 10.1109/COMST.2023.3297395.

[14] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Commun. Surveys Tuts*., vol. 18, no. 2, pp. 1153–1176, Second quarter 2016. doi: 10.1109/COMST.2015.2494502.

[15] M. Del Giudice, V. Scuotto, B. Orlando, and M. Mustilli, "Toward the human-centered approach: A revised model of individual acceptance of AI," *Hum. Resour. Manag. Rev.*, vol. 33, no. 1, Mar. 2023, Art. no. 100856. doi: 10.1016/j.hrmr.2021.100856.

[16] D. E. Denning, "An intrusion-detection model," *IEEE Trans. Softw. Eng.*, vol. 13, no. 2, pp. 222–232, 1987. doi: 10.1109/TSE.1987.232894.

[17] S. Forrest, S. A. Hofmeyr, A. Somayaji, and T. A. Longstaff, “A sense of self for Unix processes,” in *Proc. 1996 IEEE Symp. Secur. Priv.*, 1996, pp. 120–128. doi: 10.1109/SECPRI.1996.502675.

[18] E. H. Spafford and D. Zamboni, “Intrusion detection using autonomous agents,” *Comput. Netw.*, vol. 34, no. 4, pp. 547–570, 2000. doi: 10.1016/S1389-1286(00)00136-5.

[19] W. Lee, S. J. Stolfo, and K. W. Mok, “A data mining framework for building intrusion detection models,” in *Proc. 1999 IEEE Symp. Secur. Priv.*, 1999, pp. 120–132. doi: 10.1109/SECPRI.1999.766909.

[20] S. Mukkamala, G. Janoski, and A. Sung, “Intrusion detection using neural networks and support vector machines,” *Proc. 2002 Int. Joint Conf. Neural Netw.*, vol. 2, pp. 1702–1707, 2002. doi: 10.1109/IJCNN.2002.1007774.

[21] S. A. Repalle and V. R. Kolluru, “Intrusion detection system using AI and machine learning algorithm,” *Int. Res. J. Eng. Technol.*, vol. 4, no. 12, pp. 1709–1715, 2017.

[22] I. H. Abdulqadder, S. Zhou, D. Zou, I. T. Aziz, and S. M. A. Akber, “Multi-layered intrusion detection and prevention in the SDN/NFV enabled cloud of 5G networks using AI-based defense mechanisms,” *Comput. Netw.*, vol. 179, no. 4, 2020, Art. no. 107364. doi: 10.1016/j.comnet.2020.107364.

[23] S. Patil *et al.*, “Explainable artificial intelligence for intrusion detection system,” *Electronics*, vol. 11, no. 19, 2022, Art. no. 3079. doi: 10.3390/electronics11193079.

[24] A. Dutta and S. Kant, “An overview of cyber threat intelligence platform and role of artificial intelligence and machine learning,” in *Inform. Syst. Secur.: 16th Int. Conf., ICISS 2020*, Jammu, India, Springer International Publishing, Dec. 16–20, 2020, pp. 81–86.

[25] G. Sanghani and K. Kotecha, “Incremental personalized E-mail spam filter using novel TFDCR feature selection with dynamic feature update,” *Expert Syst. Appl.*, vol. 115, pp. 287–299, Jan. 2019. doi: 10.1016/j.eswa.2018.07.049.

[26] E. Cole, *Advanced Persistent Threat*. Oxford, UK: Syngres, 2012.

[27] S. M. Tahsien, H. Karimipour, and P. Spachos, “Machine learning-based solutions for security of Internet of things (IoT): A survey,” *J. Netw. Comput. Appl.*, vol. 161, Jul. 2020, Art. no. 102630. doi: 10.1016/j.jnca.2020.102630.

[28] B. Biggio and F. Roli, “Wild patterns: Ten years after the rise of adversarial machine learning,” *Pattern Recognit.*, vol. 84, no. 3, pp. 317–331, Dec. 2018. doi: 10.1016/j.patcog.2018.07.023.

[29] V. Shah, “Machine learning algorithms for cybersecurity: Detecting and preventing threats,” *Revista Espanola De Documentacion Cientifica*, vol. 15, no. 4, pp. 42–66, 2021.

[30] B. Dash, M. F. Ansari, P. Sharma, and A. Ali, “Threats and opportunities with AI-based cyber security intrusion detection: A review,” *Int. J. Softw. Eng. Appl. (IJSEA)*, vol. 13, no. 5, pp. 13–21, 2022. doi: 10.5121/ijsea.2022.13502.

[31] G. Apruzzese *et al.*, “The role of machine learning in cybersecurity,” *Digit. Threats: Res. Pract.*, vol. 4, no. 1, pp. 1–38, 2023.

[32] A. F. Jahwar and S. Y. Ameen, “A review on cybersecurity based on machine learning and deep learning algorithms,” *J. Soft Comput. Data Min.*, vol. 2, no. 2, pp. 14–25, 2021. doi: 10.30880/jscdm.2021.02.02.002.

[33] S. Samtani, M. Kantarcioglu, and H. Chen, “Trailblazing the artificial intelligence for cybersecurity discipline,” *ACM Trans. Manage. Inf. Syst.*, vol. 11, no. 4, pp. 1–19, Oct. 2020. doi: 10.1145/3430360.

[34] S. Neupane *et al.*, “Explainable intrusion detection systems (X-IDS): A survey of current methods, challenges, and opportunities,” *IEEE Access*, vol. 10, no. 7, pp. 112392–112415, 2022. doi: 10.1109/ACCESS.2022.3216617.

[35] B. Geluvaraj, P. M. Satwik, and T. A. Ashok Kumar, “The future of cybersecurity: Major role of artificial intelligence, machine learning, and deep learning in cyberspace,” in *Int. Conf. Comput. Netw. Commun. Technol.: ICCNCT 2018*, Springer Singapore, 2019, pp. 739–747.

[36] J. H. Li, “Cyber security meets artificial intelligence: A survey,” *Front. Inform. Technol. Electr. Eng.*, vol. 19, no. 12, pp. 1462–1474, 2018. doi: 10.1631/FITEE.1800573.

[37] M. Ahsan, K. Nygard, R. Gomes, M. Chowdhury, N. Rifat and J. Connolly, “Cybersecurity threats and their mitigation approaches using machine learning—A review,” *J. Cybersecurity Privacy*, vol. 2, no. 3, Sep. 2022, Art. no. 27. doi: 10.3390/jcp2030027.

[38] S. Zeadally, E. Adi, Z. Baig, and I. Khan, "Harnessing artificial intelligence capabilities to improve cybersecurity," *IEEE Access*, vol. 8, pp. 23817–23837, 2020. doi: 10.1109/ACCESS.2020.2968045.

[39] T. Stevens, "Knowledge in the grey zone: AI and cybersecurity," *Digit. War.*, vol. 1, no. 1–3, pp. 164–170, Dec. 2020. doi: 10.1057/s42984-020-00007-w.

[40] A. Lone, S. Mustajab, and M. Alam, "A comprehensive study on cybersecurity challenges and opportunities in the IoT world," *Secur. Priv.*, vol. 6, no. 6, 2023. doi: 10.1002/spy2.318.

[41] P. Donepudi, "Crossing point of artificial intelligence in cybersecurity," *Am. J. Trade Policy*, vol. 2, no. 3, pp. 121–128, Jul. 2015. doi: 10.18034/ajtp.v2i3.493.

[42] M. Musser and A. Garriott, *Machine Learning and Cybersecurity*. Washington, DC, USA: Center for Security and Emerging Technology, 2021. doi: 10.51593/2020CA004.

[43] Y. Abushark *et al.*, "Cyber security analysis and evaluation for intrusion detection systems," *Comput. Mater. Contin.*, vol. 72, no. 1, pp. 1765–1783, Jul. 2022. doi: 10.32604/cmc.2022.025604.

[44] S. Zhang, X. Xie, and X. Yang, "A brute-force black-box method to attack machine learning-based systems in cybersecurity," *IEEE Access*, vol. 8, pp. 128250–128263, 2020. doi: 10.1109/ACCESS.2020.3008433.

[45] O. M. Ijiga, I. P. Idoko, G. I. Ebiega, F. I. Olajide, T. I. Olatunde and C. Ukaegbu, "Harnessing adversarial machine learning for advanced threat detection: AI-driven strategies in cybersecurity risk assessment and fraud prevention," *Open Access Res. J. Sci. Technol.*, vol. 11, no. 1, pp. 1–24, 2024. doi: 10.53022/oarjst.2024.11.1.0060.

[46] S. Zaman *et al.*, "Security threats and artificial intelligence-based countermeasures for Internet of things networks: A comprehensive survey," *IEEE Access*, vol. 9, pp. 94668–94690, Jul. 2021. doi: 10.1109/ACCESS.2021.3089681.

[47] I. H. Sarker, "Machine learning for intelligent data analysis and automation in cybersecurity: Current and future prospects," *Ann. Data Sci.*, vol. 10, no. 6, pp. 1473–1498, 2023. doi: 10.1007/s40745-022-00444-2.

[48] T. Shojarazavi, H. Barati, and A. Barati, "A wrapper method based on a modified two-step league championship algorithm for detecting botnets in IoT environments," *Computing*, vol. 104, no. 8, pp. 1753–1774, 2022. doi: 10.1007/s00607-022-01070-9.

[49] T. Shojarazavi and A. Barati, "A survey on botnet detection methods in the Internet of things," *Int. J. Smart Electr. Eng.*, vol. 4, no. 2, pp. 99–111, 2023. doi: 10.30495/ijsee.2023.1982835.1261.

[50] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?: Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2016, pp. 1135–1144.

[51] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Int. Conf. Neural Inform. Process. Syst.*, 2017, pp. 4768–4777.

[52] K. Shaukat *et al.*, "Performance comparison and current challenges of using machine learning techniques in cybersecurity," *Energies*, vol. 13, no. 10, 2020, Art. no. 2509. doi: 10.3390/en13102509.

[53] A. Manoharan and M. Sarker, "Revolutionizing cybersecurity: Unleashing the power of artificial intelligence and machine learning for next-generation threat detection," vol. 4, no. 12, pp. 2151–2164, 2022. doi: 10.56726/IRJMETS32644.

[54] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Comput. Surveys*, vol. 54, no. 6, pp. 1–35, Jul. 2021.

[55] N. Mohamed, "Current trends in AI and ML for cybersecurity: A state-of-the-art survey," *Cogent Eng.*, vol. 10, no. 2, 2023, Art. no. 2272358. doi: 10.1080/23311916.2023.2272358.

[56] A. Barredo Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, no. 3, pp. 82–115, Jun. 2020. doi: 10.1016/j.inffus.2019.12.012.

[57] R. Sharda, D. Delen, and E. Turban, "Artificial intelligence: Concepts, drivers, major technologies, and business applications," in *Analytics, Data Science, & Artificial intelligence: Systems for Decision Support*, 11th ed. Hoboken, NJ, USA: Pearson Education, 2020, pp. 73–114.

[58] T. Q. Sun and R. Medaglia, "Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare," *Gov. Inf. Q.*, vol. 36, no. 2, pp. 368–383, Apr. 2019. doi: 10.1016/j.giq.2018.09.008.

[59] M. I. Jordan and T. M. Mitchell, “Machine learning: Trends, perspectives, and prospects,” *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015. doi: 10.1126/science.aaa8415.

[60] I. Goodfellow, Y. Bengio, and A. Courville, “Machine learning basics,” in *Deep Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 2016, vol. 1, pp. 95–151.

[61] A. Nassar and M. Kamal, “Machine learning and big data analytics for cybersecurity threat detection: A Holistic review of techniques and case studies,” *J. Artif. Intell. Mach. Learn. Manag.*, vol. 5, no. 1, pp. 51–63, Jan. 2021.

[62] A. Pumsirirat and L. Yan, “Credit card fraud detection using deep learning based on autoencoder and restricted Boltzmann machine,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 1, pp. 1–8, 2018.

[63] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward generating a new intrusion detection dataset and intrusion traffic characterization,” in *Proc. 4th Int. Conf. Inf. Syst. Security Privacy (ICISSP)*, vol. 1, pp. 108–116, 2018. doi: 10.5220/0006639801080116.

[64] J. Saxe and K. Berlin, “Deep neural network based malware detection using two dimensional binary program features,” in *2015 10th Int. Conf. Malicious Unwanted Softw. (MALWARE)*, 2015, pp. 11–20.

[65] G. R. MR, C. M. Ahmed, and A. Mathur, “Machine learning for intrusion detection in industrial control systems: Challenges and lessons from experimental evaluation,” *Cybersecurity*, vol. 4, no. 1, Dec. 2021, Art. no. 27. doi: 10.1186/s42400-021-00095-5.

[66] Q. H. Kha, V. H. Le, T. N. K. Hung, N. T. K. Nguyen, and N. Q. K. Le, “Development and validation of an explainable machine learning-based prediction model for drug-food interactions from chemical structures,” *Sensors*, vol. 23, no. 8, 2023, Art. no. 3962. doi: 10.3390/s23083962.

[67] S. Waghmare, F. Kazi, and N. Singh, “Data driven approach to attack detection in a cyber-physical smart grid system,” in *Proc. 2017 Indian Control Conf. (ICC)*, Jan. 2017, pp. 271–276.

[68] A. Saputra, “Fraud detection using machine learning in e-commerce,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 9, pp. 1–6, Sep. 2019. doi: 10.14569/issn.2156-5570.

[69] R. Shwartz-Ziv and A. Armon, “Tabular data: Deep learning is not all you need,” *Inf. Fusion*, vol. 81, no. 1, pp. 84–90, Mar. 2022. doi: 10.1016/j.inffus.2021.11.011.

[70] Y. Suseno, C. Chang, M. Hudik, and E. S. Fang, “Beliefs, anxiety and change readiness for artificial intelligence adoption among human resource managers: The moderating role of high-performance work systems,” in *Artificial Intelligence and International HRM*, Routledge, 2023, pp. 144–171.

[71] L. Floridi *et al.*, “AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations,” *Minds Mach.*, vol. 28, no. 4, pp. 689–707, Nov. 2018. doi: 10.1007/s11023-018-9482-5.

[72] B. T. Familoni, “Cybersecurity challenges in the age of AI: Theoretical approaches and practical solutions,” *Comput. Sci. IT Res. J.*, vol. 5, no. 3, pp. 703–724, 2024.

[73] F. Siddiqui, R. Khan, and S. Sezer, “Bird’s-eye view on the Automotive Cybersecurity Landscape & Challenges in adopting AI/ML,” in *2021 Sixth Int. Conf. Fog Mob. Edge Comput. (FMEC)*, IEEE, 2021, pp. 1–6.

[74] G. S. Nadella and H. Gonaygunta, “Enhancing cybersecurity with artificial intelligence: Predictive techniques and challenges in the age of IoT,” *Int. J. Sci. Eng. Appl.*, vol. 13, no. 4, pp. 30–33, 2024.

[75] A. D. Sontan and S. V. Samuel, “The intersection of Artificial Intelligence and cybersecurity: Challenges and opportunities,” *World J. Adv. Res. Rev.*, vol. 21, no. 2, pp. 1720–1736, 2024. doi: 10.30574/wjarr.2024.21.2.0607.

[76] R. Xu, N. Baracaldo, and J. Joshi, “Privacy-preserving machine learning: Methods, challenges and directions,” Aug. 2021, *arXiv:2108.04417*.

[77] M. F. Ansari, B. Dash, P. Sharma, and N. Yathiraju, “The impact and limitations of artificial intelligence in cybersecurity: A literature review,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 11, no. 9, pp. 81–90, 2022. doi: 10.17148/IJARCCE.2022.11912.

[78] M. Macas, C. Wu, and W. Fuertes, “A survey on deep learning for cybersecurity: Progress, challenges, and opportunities,” *Comput. Netw.*, vol. 212, no. 3, 2022, Art. no. 109032. doi: 10.1016/j.comnet.2022.109032.

[79] J. Qin, B. Liu, and J. Qian, "A novel privacy-preserved recommender system framework based on federated learning," in *Proc. 2021 4th Int. Conf. Softw. Eng. Inf. Manag.*, Jan. 2021, pp. 82–88.

[80] A. Handa, A. Sharma, and S. K. Shukla, "Machine learning in cybersecurity: A review," *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.*, vol. 9, no. 4, Jul. 2019, Art. no. e1306.

[81] J. P. Bharadiya, "AI-driven security: How machine learning will shape the future of cybersecurity and Web 3.0," *Am. J. Neural Netw. Appl.*, vol. 9, no. 1, pp. 1–7, 2023. doi: 10.11648/j.ajnna.20230901.11.

[82] A. P. Veiga, "Applications of artificial intelligence to network security," 2018, *arXiv:1803.09992*.

[83] I. Kotenko, I. Saenko, and A. Branitskiy, "Machine learning and big data processing for cybersecurity data analysis," in *Data Science in Cybersecurity and Cyberthreat Intelligence*, Cham: Springer, 2020, pp. 61–85.

[84] R. Raimundo and A. Rosário, "The impact of artificial intelligence on data system security: A literature review," *Sensors*, vol. 21, no. 21, 2021, Art. no. 7029. doi: 10.3390/s21217029.

[85] E. Bertino, M. Kantarcioglu, C. G. Akcora, S. Samtani, S. Mittal and M. Gupta, "AI for security and security for AI," in *Proc. Eleventh ACM Conf. Data Appl. Secur. Priv.*, Apr. 2021, pp. 333–334.

[86] R. Kaur, D. Gabrijelčič, and T. Klobučar, "Artificial intelligence for cybersecurity: Literature review and future research directions," *Inf. Fusion*, vol. 97, no. 6, 2023, Art. no. 101804. doi: 10.1016/j.inffus.2023.101804.

[87] R. Khan, P. Kumar, D. N. K. Jayakody, and M. Liyanage, "A survey on security and privacy of 5G technologies: Potential solutions, recent advancements, and future directions," *IEEE Commun. Surv. Tutor.*, vol. 22, no. 1, pp. 196–248, 2019. doi: 10.1109/COMST.2019.2933899.

[88] G. N. Brijwani, P. E. Ajmire, and P. V. Thawani, "Future of quantum computing in cybersecurity," in *Handbook of Research on Quantum Computing for Smart Environments*. Hershey, PA, USA: IGI Global, 2023, pp. 267–298.

[89] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, Jan. 2019. doi: 10.1145/3339474.

[90] S. Al-Mansoori and M. B. Salem, "The role of artificial intelligence and machine learning in shaping the future of cybersecurity: Trends, applications, and ethical considerations," *Int. J. Soc. Analy.*, vol. 8, no. 9, pp. 1–16, 2023.

[91] C. Okoli and K. Schabram, "A guide to conducting a systematic literature review of information systems research," *Commun. Assoc. Inform. Syst.*, vol. 37, no. 1, pp. 1–66, 2010.

[92] M. -I. Mahraz, L. Benabbou, and A. Berrado, "A systematic literature review of digital transformation," in *Proc. Int. Conf. Indus. Eng. Operat. Manag.*, Toronto, ON, Canada, IEOM Society International, Oct. 2019, pp. 23–25.

[93] A. Shukla, "Leveraging AI and ML for advance cyber security," *J. Artif. Intell. Cloud Comput.*, vol. 1, no. 1, 2022, Art. no. 142. doi: 10.47363/JAICC/2022(1)142.

[94] A. Bécue, I. Praça, and J. Gama, "Artificial intelligence, cyber-threats and Industry 4.0: Challenges and opportunities," *Artif. Intell. Rev.*, vol. 54, no. 5, pp. 3849–3886, 2021. doi: 10.1007/s10462-020-09942-2.

[95] J. Kinyua and L. Awuah, "AI/ML in security orchestration, automation and response: Future research directions," *Int. Autom. Soft Comput.*, vol. 28, no. 2, pp. 527–545, 2021. doi: 10.32604/iasc.2021.016240.