Journal of Knowledge Learning and Science Technology
ISSN: 2959-6386 (Online)
2024, Vol. 4, No. 1, pp. 103–111
DOI: https://doi.org/10.60087/jklst.v4.n1.011

Research Article

A Review on the Effectiveness of Artificial Intelligence and Machine Learning on Cybersecurity

Ojo, Abraham O.

Publicis Sapient, Houston, TX 77002, United States of America

ORCID

https://orcid.org/0009-0005-9906-2854

Abstract

The rapid growth of cyber threats and increasing complexity of attack techniques demand advanced solutions for protecting systems, networks, and sensitive data. Artificial Intelligence (AI) and Machine Learning (ML) have proven to be highly effective in a wide range of cybersecurity applications, ranging from detecting malicious activities to automating defense mechanisms. These technologies are transforming the design and operation of security systems. This has become essential for addressing the evolving landscape of digital threats. With the exponential rise in cyberattacks and the sophistication of adversaries, traditional cybersecurity methods are proving inadequate. AI and ML offer innovative solutions by enhancing threat detection, improving response times, and automating tasks requiring significant human intervention. These technologies are widely employed in applications such as anomaly detection, intrusion detection systems (IDS), malware analysis, and fraud detection. The primary benefits of AI and ML in cybersecurity include real-time threat analysis, predictive capabilities, and ability to process large volumes of data. However, they also present challenges such as high implementation costs, the need for skilled professionals, and the potential for adversarial attacks that can exploit machine learning models. This review aims to provide a comprehensive understanding of the integration of AI and ML into cybersecurity, highlighting their current applications, benefits, limitations, and prospects. Key findings indicate that, while these technologies significantly enhance cybersecurity capabilities, their success depends on overcoming challenges related to data quality, model robustness, and the dynamic nature of cyber threats.

Keywords

Artificial intelligence, machine learning, cybersecurity, ransomware

Introduction

The digital era has ushered in an unprecedented level of connectivity, data sharing, and reliance on technology, all of which have revolutionized business operations, communication, and social interactions. However, these advancements have expanded the attack surface for malicious actors, making cybersecurity one of the most critical concerns in today's world (Kaur, Gabrijelcic and Klobucar, 2023). Cybersecurity challenges have evolved rapidly, requiring innovative and

*Corresponding author: Ojo, Abraham O.

Email addresses:

abraham.cyberdefender@gmail.com

Received: 01-11-2024; **Accepted:** 01-12-2024; **Published:** 25-01-2025

sophisticated solutions to safeguard against the growing threat landscape; hence, the integration of Artificial Intelligence and Machine Learning into cybersecurity strategies has emerged as a promising way to enhance threat detection, response, and prevention (Musa et al., 2024). This review delves into the significance of AI and ML in cybersecurity and explores their effectiveness in mitigating the challenges posed by modern cyber threats.

The digital transformation of businesses and proliferation of connected devices have brought forth a host of new cybersecurity challenges. These challenges are compounded by the increasing sophistication of cybercriminals, rise of advanced persistent threats (APTs), and rapid evolution of cyberattack techniques (Li et al., 2022). In terms of the complexity of threats, cyberattacks today are more varied, complex, and persistent than ever before, and such threats include ransomware, phishing attacks, and data breaches, which emerge frequently, thereby making it difficult for traditional defense mechanisms to keep up (Torres, Comesana and García-Nieto, 2019). The exponential growth in the data generated by users, devices, and organizations adds another layer of complexity to cybersecurity. That is, the sheer volume of data often overwhelms traditional security systems, which struggle to process and analyze large datasets in real time (Raimundo and Rosario, 2021). With respect to the speed of attack, modern cyberattacks are executed with incredible speed, leaving minimal time for human responders to detect and mitigate them. For instance, a zero-day vulnerability can be exploited before defenders have a chance to patch it (Redino et al., 2021). Similarly, internal actors, either malicious or negligent, pose a significant security risk, and these insider threats are harder to detect, as they often involve authorized access to systems and data (Rauf, Mohsen and Wei, 2023). The role of human error cannot be overemphasized because, despite advancements in technology, human error continues to be a primary factor in cybersecurity incidents, as phishing emails, weak passwords, and improper configurations remain common entry points for cybercriminals (Hansen and Bøgh, 2021). From the perspective of adversarial attacks, the growing use of AI and ML by attackers also presents a new challenge, as these technologies can be used to launch attacks that are specifically designed to evade detection by traditional systems, thus requiring organizations to explore innovative approaches to enhance their cybersecurity capabilities (Fritsch, Jaber and Yazidi, 2022; Reilly, O'Shaughnessy and Thorpe, 2023). Traditional methods, although effective to some extent, often fail to address the dynamic and increasingly complex nature of cyber threats.

Artificial Intelligence (AI) and Machine Learning (ML) in Cybersecurity

Artificial Intelligence and Machine Learning represent a class of technologies that have demonstrated tremendous potential in automating tasks, solving complex problems, and analyzing vast amounts of data. In the context of cybersecurity, AI refers to the simulation of human intelligence processes by machines, enabling them to learn from data, recognize patterns, and make decisions, whereas Machine Learning, a subset of AI, involves training algorithms to identify patterns in data and use this information to make predictions or decisions without explicit programming (Kaur et al., 2023; Jada and Mayayise, 2024). In cybersecurity, AI and ML are explored in various ways to strengthen security defenses and improve response mechanisms. Some key applications of AI and ML in cybersecurity include the following.

Threat Detection and Prediction: According to Wazid et al. (2022), AI and ML algorithms can analyze large volumes of network traffic, user behavior, and system logs to detect anomalies and predict potential security incidents. For instance, machine learning models can be trained to recognize patterns of normal network traffic and identify deviations that may signal an attack, such as a DDoS attack or a data-exfiltration attempt.

Automated Incident Response: AI and ML can be used to automate certain aspects of incident response, such as identifying the root cause of an attack, containing the threat, and deploying countermeasures (Salem et al., 2024). This reduces the time required to respond to incidents and minimizes the damage caused by cyberattacks.

Malware Detection: AI and ML are instrumental in identifying new and evolving malware by recognizing patterns in behavior, rather than relying solely on signature-based detection methods, which helps detect previously unknown threats, such as zero-day attacks, more effectively (Muhamed, 2023). Malware analysis is a critical aspect of cybersecurity because understanding the behavior and structure of malicious software is essential for detecting, neutralizing, and preventing attacks. In terms of automated malware identification and classification, traditional malware analysis techniques often rely on human analysts who manually examine files and identify known malware signatures (Truong et al., 2020). However, this process is time consuming, error prone, and ineffective against new or polymorphic malware. AI and ML enhance malware analysis by automating the identification and classification of malicious software because ML models can be trained on large datasets of known malware samples to identify features such as file structure, behavior patterns, or code signatures, which are indicative of malicious activity (Vanjire and Lakshmi, 2021). These models can then classify new files as benign or malicious based on their similarities to previously seen samples. One of the most challenging types of malwares to detect is polymorphic malware, which changes its code or behavior to evade detection using traditional antivirus software (Okdem and Okdem, 2024). Obfuscated malware, on the other hand, hides its true intent by encrypting or disguising its

payload (Shafin, Karmakar and Mareels, 2023). Machine learning plays a critical role in identifying these evasive threats because its algorithms can detect patterns in the behavior of obfuscated or polymorphic malware by analyzing the underlying activity, even when the code itself has been altered. For example, an ML model might recognize that a file repeatedly attempts to connect to a suspicious IP address regardless of the specific code used to execute the attack (Vanjire and Lakshmi, 2021; Shafin, Karmakar and Mareels, 2023). ML models can effectively identify and classify obfuscated and polymorphic malware even when and/or where traditional signature-based methods fail.

Fraud Detection: Fraud detection is a critical application of AI and ML, particularly in the financial sector, where fraudulent transactions can result in significant financial losses (Ghandour, 2021). Traditional fraud-detection systems rely on rule-based algorithms, which are often unable to keep up with the complexity and sophistication of modern fraudulent activities. However, AI and ML improve fraud detection by analyzing transaction patterns and user behavior to identify anomalies that could indicate fraudulent activity (Torres et al., 2019). Machine learning is widely used in financial services and e-commerce to detect fraudulent activities (Welukar and Bajoria, 2021). ML can be used to analyze transaction patterns and user behavior, equally flag any suspicious activity, and prevent financial fraud or identity theft. Machine learning models can be trained on historical transaction data to recognize patterns associated with legitimate transactions and to identify deviations that may signal fraud. The main applications include credit card fraud detection, in which ML algorithms can analyze transaction data in real time, looking for unusual spending patterns or geographic locations that may indicate fraudulent activity (Welukar and Bajoria, 2021). If a user typically makes transactions in one location, but a transaction is detected from another country, the system can flag this as potentially fraudulent because it can detect account takeovers, unauthorized fund transfers, and other types of online banking fraud by analyzing login patterns, device identifiers, and transaction histories (Ma, Dhot and Raza, 2023). Machine learning models can learn the normal behavior of users and flag unusual activities for further investigation. In e-commerce, AI can also detect fraudulent activities, such as payment fraud, fake reviews, or account creation fraud, by analyzing user behavior, IP addresses, payment methods, and browsing patterns, making ML and AI versatile and highly adaptive to real-time threats and other evolving fraud tactics (Kant and Johannsen, 2022).

Phishing Detection Phishing attacks are among the most common and damaging forms of cybercrime. Attackers use deceptive emails, websites, and messages to trick individuals to reveal sensitive information such as passwords, credit card numbers, or personal details (Torres et al., 2019; Hansen and Bøgh, 2021). Traditional phishing detection methods typically rely on manual rules or keyword filtering, which are often insufficient because of the constantly evolving tactics used by attackers, whereas AI and ML techniques can significantly improve phishing detection in both emails and websites by analyzing various factors and identifying subtle patterns indicative of malicious activity (Indrasiri et al., 2021). Specifically, ML algorithms can analyze email content to identify potential phishing attempts by examining factors such as suspicious or malformed email addresses, unusual languages or urgent tones that are characteristic of phishing attempts, malicious attachments or links, and inconsistencies in the sender's domain or content that may suggest spoofing (Truong et al., 2020). For the website verification, phishing websites are designed to mimic legitimate sites, often tricking users into entering their credentials owing to the fact that ML models can analyze websites to identify signs of phishing, such as, suspicious URLs or domains that mimic trusted brands, SSL certificate anomalies or the absence of HTTPS, and suspicious design elements or content inconsistencies (Shafin et al., 2023). Through this verification, the legitimacy of websites in real time can prevent users from accessing malicious websites, reducing the risk of credential theft and fraud. According to Torres et al. (2019), ML algorithms can analyze emails, websites, and other online communications to identify the characteristics of phishing attempts, which can be used to detect subtle patterns that humans might miss, such as deceptive languages or malicious URLs. From all the indications, it is apparent that AI and ML enhance traditional cybersecurity techniques by providing faster, more accurate, and scalable solutions for identifying and mitigating threats. However, their implementation also introduces new challenges, including the risk of adversarial attacks against machine learning models, data privacy concerns, and the need for continuous model retraining. The effectiveness of AI and ML in cybersecurity is a topic of increasing interest because of the potential to revolutionize how organizations defend themselves against cyber threats. As organizations integrate AI and ML technologies into their cybersecurity frameworks, it is essential to assess their actual performance in real-world environments and to understand their limitations and potential risks.

The significance of studying the effectiveness of AI and ML lies in the need to evaluate their real-world effectiveness in addressing the rapidly evolving cyber threat landscape, which includes assessing their accuracy, efficiency, and scalability in detecting and mitigating attacks (Fritsch et al., 2022). Although AI and ML offer significant advantages, it is also crucial to understand their limitations, such as susceptibility to adversarial attacks, data quality issues, and the requirement for continuous training, as emphasized by Tan et al. (2020). Similarly, looking at the effectiveness of AI and ML allows researchers and experts to identify areas for improvement, develop more robust algorithms, enhance model interpretability, ensure privacy protection, guide future advancements in these

technologies, and ensure that they remain relevant and capable of addressing emerging threats (Indrasiri, Halgamuge, and Mohammad, 2021).

Artificial Intelligence and Machine Learning have become transformative forces in modern cybersecurity, offering powerful solutions for detecting, predicting, and responding to an increasingly sophisticated range of cyber threats (Salem et al., 2024). As cyberattacks evolve in complexity and scale, traditional cybersecurity methods that rely heavily on human intervention, signature-based systems, and predefined rules are becoming less effective in addressing these dynamic threats. AI and ML, with their ability to process vast amounts of data, learn from it, and make autonomous decisions, offer significant advantages for enhancing cybersecurity (Li et al., 2022).

Characteristics of AI and ML in the Context of Cybersecurity:

In context of cybersecurity, AI and ML systems are capable of making decisions and taking actions independently without human intervention. For example, an AI-driven system can automatically identify and block potential threats in real time without waiting for manual analysis (Haider et al., 2020). Equally, both AI and ML rely on large datasets to learn and generate insights, and this data-centric approach allows them to analyze and detect previously unseen threats based on patterns and behaviors, rather than relying solely on predefined rules or signatures (Kaur et al., 2023). In addition, ML models continuously learn and adapt to new patterns that enable them to detect evolving threats, such as zero-day attacks or novel malware, which may not be recognized by traditional signature-based systems (Redino et al., 2021). AI and ML systems are also designed to analyze vast amounts of data in real time, making them capable of quickly detecting and responding to threats, something that traditional cybersecurity tools may struggle with owing to time constraints (Wazid et al., 2022). All of these features make AI and ML particularly effective in modern cybersecurity environments, where the volume, variety, and velocity of cyber threats have grown exponentially.

Capabilities of AI/ML in Cybersecurity: The integration of AI and ML in cybersecurity offers a range of powerful capabilities that address both the complexity of modern cyberattacks and limitations of traditional defense mechanisms. Such capabilities include threat detection and response, as AI and ML excel at detecting threats in real time by analyzing data from various sources, such as network traffic, system logs, and user behavior. Traditional threat detection methods often rely on predefined rules and signatures to identify known threats, making them less effective in detecting novel or unknown attacks (Ahmad et al., 2022). In contrast, AI and ML use advanced algorithms to identify patterns and anomalies that could indicate potential threats. One of the most significant advantages of AI and ML in cybersecurity is their ability to predict potential threats before they occur (Hasan, 2019). With the analysis of historical data, such as attack patterns, malware signatures, and vulnerabilities, AI and ML models can identify indicators of compromise (IoCs) and predict where future attacks might occur (Wazid et al., 2022); which allows organizations to take proactive measures to defend themselves against threats. AI and ML are also crucial in automating various cybersecurity tasks, which is particularly important given the volume of security events that need to be processed, as automation helps reduce the workload on cybersecurity professionals and ensures that responses to threats are faster and more efficient (Jada and Mayayise, 2024). Through automation, AI and ML free up security teams to focus on more strategic tasks, while ensuring that routine security processes are handled efficiently and effectively.

Success Stories and Case Studies Demonstrating Effectiveness of AI and ML in Cybersecurity

The adoption of AI and ML in cybersecurity is rapidly expanding across various industries, from financial services to the healthcare and government sectors. Several organizations have experienced tangible outcomes from integrating AI and ML into their cybersecurity strategies (Fritsch et al., 2022). Several success stories and case studies have hitherto highlighted the transformative impact of AI and ML in cybersecurity, thereby illustrating how these technologies enhance threat detection, improve response times, and secure critical infrastructure. For instance, Darktrace is one of the leading companies in AI-powered cybersecurity and uses ML and AI algorithms to detect and respond to cybersecurity threats in real time (Mohammed, 2020). Its self-learning AI, known as the "Enterprise Immune System," uses unsupervised machine learning to understand normal patterns of behavior in a network and then detects deviations from those patterns. DarkTrace has been widely adopted across various industries, including finance, healthcare, and energy, where it has successfully detected advanced persistent threats (APTs) and insider threats that traditional signature-based systems miss. A notable case study was the event in 2020, where Darktrace was able to detect a sophisticated cyberattack targeting a major university in the UK. The attack involved a malicious insider attempting to exfiltrate sensitive academic research data (Welukar and Bajoria, 2021). Darktrace's AI system detected abnormal behavior, flagged it for immediate investigation, and helped the university prevent significant data breaches. This success demonstrates the power of AI to detect and mitigate threats that might otherwise go undetected by traditional

systems. In addition, Google's Chronicle security platform uses AI and ML to enhance threat detection and incident response (Zhang et al., 2022). By processing massive amounts of security data, the Chronicle enables organizations to identify potential security incidents and provide deep insights into the context of those events. Its ability to correlate data from different sources, such as firewalls, endpoint devices, and cloud services, allows security teams to identify complex attack patterns that may go unnoticed using traditional approaches (Zhang et al., 2022). As a case study, most major financial institutions leveraged the Chronicle's AI-driven approach to detect and respond to an internal data breach. The system automatically identifies a suspicious pattern of data access by an employee and alerts the security team, enabling them to take swift action and prevent further damage. This highlights the ability of AI-powered systems to provide more comprehensive threat visibility and accelerate response times.

Similarly, IBM Watson uses AI and cognitive computing to enhance threat detection, incident response, and overall cybersecurity operations. Watson analyzes large volumes of security data, correlates them with threat intelligence feeds, and uses natural language processing to provide security analysts with relevant insights for decision-making (Srivani, Abirami and Mala, 2023). In a 2017 collaboration with a financial services company, Watson assisted in identifying phishing attacks that were disguised as legitimate communication from banking partners. The system's ability to process and analyze text content and flag unusual patterns allowed the financial institution to stop phishing attempts before they could impact customers (Chen, Elenee and Weber, 2016). In other words, IBM Watson's ability to leverage AI to analyze complex data and detect new threats shows its effectiveness in real-time cybersecurity defense.

In healthcare, AI and ML are used to protect patient data from cyberattacks and secure healthcare networks. According to Jadav et al. (2023), machine learning models are deployed to detect unusual behaviors in hospital networks, monitor medical devices for vulnerabilities, and identify ransomware threats that target healthcare organizations. For instance, a leading hospital system in the U.S. implemented AI-driven threat detection tools to identify and prevent ransomware attacks (Musbahi et al., 2021). The system detected abnormal access patterns in the hospital network, allowing the security team to stop an attempted ransomware attack before it could encrypt critical patient data (Ahamed et al., 2022). In addition, government agencies, especially in the defense and intelligence sectors, leverage AI and ML to secure sensitive data and defend against state-sponsored cyberattacks (Cam, 2019), AI-driven systems are also applied to monitor national infrastructure, detect cyber-espionage, and identify signs of cyber warfare (Shaukat et al., 2020). The U.S. Department of Homeland Security (DHS) has adopted AI-driven security tools to monitor critical infrastructure in real time, and these tools have successfully identified and neutralized a number of cyber-espionage attempts, strengthening national security.

These success stories underscore the ability of AI and ML to proactively identify and respond to complex and evolving threats, showcasing the significant potential of these technologies in enhancing cybersecurity efforts.

Challenges and Limitations

Although Artificial Intelligence and Machine Learning offer substantial benefits for cybersecurity, their integration is not without challenges and limitations. These technologies face various obstacles, ranging from adversarial attacks to ethical concerns, which must be addressed to maximize their potential for combating cyber threats. Part of the challenges that AI and ML encounter in cybersecurity is known as "Adversarial Machine Learning" which has to do with attacks on AI/ML Models. Adversarial machine learning refers to the techniques used by attackers to manipulate or deceive AI and ML models to make incorrect predictions or classifications (Reilly et al., 2023). In cybersecurity, adversarial attacks can be highly detrimental because they can bypass detection systems or cause misclassifications of malicious behavior as benign. This can take the form of data poisoning, evasion attacks, etc. (Fritsch et al., 2022). Data poisoning occurs when attackers inject misleading or malicious data into the training set of an ML model, compromising its ability to learn and classify threats (Salem et al., 2024). For example, an attacker might inject fraudulent transactions into a financial fraud detection system, causing the system to misclassify legitimate transactions as fraudulent or vice versa. In the case of evasion attacks, the attacker subtly modifies the input data to trick the model into classifying malicious activity as harmless. For example, an attacker may modify the structure of malware code to evade detection by an ML-based malware analysis system (Shroff et al., 2022). There are also data-related challenges related to the quality, quantity, and labeling of the training data (Okdem and Okdem, 2024). It is worth stressing that the AI and ML models rely heavily on high-quality labeled data for training and validation. In cybersecurity, obtaining comprehensive and representative datasets is often challenging because of data quality, quantity, and labeling of data (Okdem and Okdem, 2024). In other words, incomplete, noisy, or inaccurate data can hinder the model's ability to learn effectively, leading to poor performance and misclassifications because cybersecurity data often require accurate labeling to ensure the correct classification of threats, whereas manual labeling can be time-consuming and error-prone, especially when dealing with complex attack scenarios (Okdem and Okdem, 2024). In addition, training models with large volumes of diverse data is necessary for generalization. However, acquiring such large datasets, especially for rare or new types of cyber threats, can be difficult (Wazid

et al., 2022).

Another significant challenge is privacy concerns regarding the data collection and usage. Security systems often analyze sensitive personal data, such as online activities or financial transactions, which can lead to privacy violations if not handled correctly (Romsom, 2022). Compliance with regulations and laws such as the General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA) imposes strict rules on how personal data must be collected, stored, and used (GDPR, 2016). Therefore, balancing the data collection for AI training with privacy regulations is a significant challenge. Furthermore, there are computational and resource constraints because training large-scale AI and ML models in cybersecurity requires significant computational power, which can be expensive and resource intensive. Large ML models also require advanced infrastructure such as high-performance GPUs or distributed computing systems to process vast datasets and perform complex computations (Wazid et al., 2022). Training AI models, particularly deep learning models, consumes significant energy, contributing to high operational costs and environmental impacts. The high costs and infrastructure requirements for large-scale ML models also constitute a major issue because the cost of developing, maintaining, and scaling AI/ML-based cybersecurity systems can be prohibitive for small- or medium-sized organizations, potentially limiting widespread adoption (Wilkins, 2018). Thus far, ensuring that AI and ML systems effectively detect and mitigate threats while respecting user privacy is a delicate balance, as the use of AI in security operations should be transparent and accountable and users should also have control over their data.

Conclusion

In conclusion, the integration of Artificial Intelligence and Machine Learning into cybersecurity has revolutionized the cybersecurity landscape and has proven to be highly effective in addressing the increasingly complex and sophisticated cyber threats that organizations face today. These technologies have significantly enhanced the detection, prevention, and mitigation of cyberattacks, enabling faster responses, more accurate threat identification, and proactive approaches to security. The ability of AI/ML to process large volumes of data, identify patterns, and adapt to evolving threats makes it a crucial asset in modern cybersecurity frameworks. However, despite their promising capabilities, notable challenges and limitations must be addressed. Adversarial attacks, data privacy concerns, and resource-intensive nature of AI/ML models remain significant obstacles. Furthermore, issues such as the need for high-quality labeled data, biases in algorithms, and regulatory compliance highlight the complexity of implementing AI/ML in cybersecurity in a way that is both effective and ethical. Although these challenges are substantial, they also present opportunities for continued innovation, particularly in developing more robust and resilient models, enhancing data privacy mechanisms, and improving the interpretability of AI-driven systems.

Recommendations

To effectively leverage AI and ML in cybersecurity, organizations must follow strategic practices that ensure that systems are deployed, maintained, and updated appropriately. The following recommendations outline the best practices for the successful implementation of AI/ML in cybersecurity.

AI/ML should be seamlessly integrated with traditional security measures (e.g., firewalls and intrusion detection systems). This ensures layered defense, where AI augments human expertise and traditional methods to improve overall security effectiveness.

The AI/ML models depend on high-quality, accurate, and diverse datasets. Ensuring the quality of training data, protecting it from tampering, and maintaining privacy standards are crucial.

Establish clear goals for AI/ML applications in cybersecurity, such as improving the threat detection accuracy and reducing incident response times. Monitor performance using key performance indicators (KPIs) to measure success and identify areas for improvement.

Cyberthreats are dynamic and continuously evolving. AI/ML models must undergo regular retraining to adapt to new attack techniques, patterns, and vulnerabilities. Automated retraining pipelines ensure that the models are relevant and effective.

The continuous collection of new labeled data is essential for training and refining AI/ML models. This ensures that the models remain effective in detecting the latest threats and can respond to emerging attack vectors.

Cybersecurity threats are becoming increasingly complex and global, making it important for organizations to collaborate across sectors. Sharing threat intelligence, attack patterns, and defense strategies can help to create more resilient defenses against cyberattacks.

Establishing trusted partnerships with other organizations, government agencies, and cybersecurity firms allows for the faster identification and mitigation of threats. This collective intelligence enhances the effectiveness of AI/ML-based systems across the cybersecurity landscapes.

As AI and ML become integral to cybersecurity, regulatory frameworks are essential for ensuring ethical and legal use. Clear guidelines should govern how data are collected, processed, and shared, with emphasis on privacy, fairness, and accountability.

Regulations should ensure that AI systems used in cybersecurity are transparent and auditable, with mechanisms for addressing issues such as bias and discrimination in decision making. Establishing standards for explainability (XAI) is

particularly important for ensuring trust in AI-driven security measures.

Funding:

This research received no funding from any source.

Ethical Considerations:

This study was carried out in accordance with the Declaration of Helsinki on research guidelines, which are: “anonymity, informed consent, privacy, confidentiality, and professionalism.”

Conflict of interest:

The authors declare that they have no conflict of interest.

References

[1] Ahamed, F., Farid, F., Suleiman, B., Jan, Z., Wahsheh, L.A., & Shahrestani, S. (2022). An intelligent Multimodal Biometric Authentication model for Personalised healthcare services. *Future Internet,* 14(8). https://doi.org/10.3390/fi14080222

[2] Cam, H. (2019). Model-guided infection prediction and active defense using context-specific cybersecurity observations at the MILCOM 2019-2019 IEEE Military Communications Conference (MILCOM). pp. 1-6.

[3] Chen, Y., Elenee, A. J. D., & Weber, G. (2016). IBM Watson: How Cognitive Computing Can be Applied to Big Data Challenges in Life Sciences Research. *Clin Ther*., 38(4), 688-701. https://doi.org/10.1016/j.clinthera.2015.12.001

[4] Fritsch, L., Jaber, A., & Yazidi, A. (2022). Overview of artificial intelligence used in malware. In *Communications in computer and information science, 1650 CCIS*. https://doi.org/10.1007/978-3-031-17030-0_4

[5] Gawand, M.K. (2013). Comparative study of cyber-attack detection and prediction using machine learning algorithms. https://doi.org/10.21203/rs.3.rs- 32385 52/v1

[6] General Data Protection Regulation (GDPR), (2016). "General Data Protection Regulation," European Union, Brussels, Belgium, Regulation 2016/679.

[7] Ghandour, A. (2021). Opportunities and challenges of artificial intelligence in banking: a systematic literature review. *TEM Journal, 10*(4), 1581-1587. https://doi.org/10.18421/TEM104-12

[8] Haider, A., Khan, M.A., Rehman, A., Ur-Rahman, M., & Kim, H.S. (2020). Real-time sequential deep extreme learning machine cybersecurity intrusion detection system. *Computers, Materials & Continua,* 66(2), 1785-1798. https://doi.org/10.32604/cmc.2020.013910

[9] Hansen, E.B., & Bøgh, S. (2021). Artificial intelligence and Internet of Things in small- and medium-sized enterprises: A survey. *Journal of Manufacturing Systems,* 58, 362-372.

[10] Hasan, M., Islam, M.M., Zarif, M.I.I., Hashem, M.M.A. (2019). Attack and anomaly detection in IoT sensors at IoT sites using machine learning approaches. *Internet Things*, 7, 100059. https://doi.org/10.1016/j.iot. 2019.100059

[11] Indrasiri, P.L., Halgamuge, M.N., Mohammad, A. (2021). Robust ensemble machine learning model for filtering phishing URLs: expandable random gradient stacked voting classifier (ERG-SVC), *IEEE Access*, 9, 150142–150161.

[12] Jada, I., & Mayayise, T.O. (2024). Impact of artificial intelligence on organizational cyber security: An outcome of a systematic literature review. *Data and Information Management*, 8, 100063. https://doi.org/10.1016/j.dim.2023.100063

[13] Jadav, D., Jadav, N. K., Gupta, R., Tanwar, S., Alfarraj, O., Tolba, A., et al. (2023). A trustworthy healthcare management framework using the amalgamation of AI and a blockchain network. *Mathematics, 11*(3). https://doi.org/10.3390/math11030637

[14] Kant, D. and Johannsen, A. (2022). Evaluation of AI-based use cases to enhance cybersecurity defense of small and medium-sized companies (SMEs). *IS and T International Symposium on Electronic Imaging Science and Technology,* 34(3). https://doi.org/10.2352/EI.2022.34.3.MOBMU-387

[15] Kaur, R., Gabrijelcic, D., & Klobucar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future directions. *Information Fusion* 97, 101804. https://doi.org/10.1016/j.inffus.2023.101804

[16] Li, H., Wu, J., Xu, H., Li, G., & Guizani, M. (2022). Explainable intelligence-driven defense mechanism against advanced persistent threats: A joint edge game and an AI approach. *IEEE Trans. Dependable Secure Comput*. 19(2), 757-775.

[17] Ma, K.W.F., Dhot, T., & Raza, M. (2023). Considerations for using artificial intelligence to manage authorized push payment (APP) scams. In *IEEE engineering management review*. https://doi.org/10.1109/EMR.2023.3288432

[18] Mohamed, N. (2023). Current trends in AI and ML for cybersecurity: A state-of-the-art survey. *Cogent Eng*., https://doi.org/10.1080/23311916.2023.2272358

[19] Mohammed, I.A. (2020). Artificial intelligence for cybersecurity: a systematic mapping of literature. *International Journal of Innovations in Engineering Research and Technology*, 7(9), 2394-3696.

[20] Musa, N.S., Mirza, N.M., Rafique, S.H., Abdallah, A.M., Murugan, T. (2024). Machine learning and deep learning

techniques for distributed denial of service anomaly detection in software-defined networks are current research solutions. *IEEE Access*, 12. https://doi.org/10.1109/ACCESS.2024.3360868

[21] Musbahi, O., Syed, L., Le Feuvre, P., Cobb, J., Jones, G. (2021). Public patient views on artificial intelligence in healthcare: A nominal group technique study. *Digital Health,* 7. https://doi.org/10.1177/20552076211063682

[22] Okdem, S. and Okdem, S. (2024). Artificial Intelligence in Cybersecurity: A Review and a Case Study. *Appl. Sci*., 14, 10487. https://doi.org/10.3390/app142210487

[23] Raimundo, R., & Rosario, A. (2021). Impact of artificial intelligence on Data System Security: A Literature Review. *Sensors,* 21(21), 7029.

[24] Rauf, U., Mohsen, F., Wei, Z. (2023). Taxonomic classification of insider threats: Existing techniques, future directions, and recommendations. *J. Cyber Secur Mobil*., 12(2), 221–252. https://doi.org/10.13052/jcsm2245-1439.1225

[25] Redino, C., Nandakumar, D., Schiller, R., Choi, K., Rahman, A., Bowen, E. et al. (2022). Zero-day threat detection using graph- and flow-based security telemetry. *3rd IEEE 2022 International Conference on Computing, Communication, and Intelligent Systems, ICCCIS,* 655–662. https://doi.org/10.1109/ICCCIS56430.2022.10037596

[26] Reilly, C., O'Shaughnessy, S., Thorpe, C. (2023). Robustness of image-based malware classification models trained using generative adversarial networks. *ACM International Conference Proceeding Series* 92–99. https://doi.org/10.1145/3590777.3590792

[27] Romsom, E. (2022). *Countering global oil theft: Responses and solutions*: WIDER Working Paper.

[28] Salem, A.H., Azzam, S.M., Emam, O.E., Abohany, A.A. (2024). Advancement Cybersecurity: A comprehensive review of AI-driven techniques. *Journal of Big Data* 11(105). https://doi.org/10.1186/s40537-024-00957-y

[29] Shafin, S. S., Karmakar, G., Mareels, I. (2023). Obfuscated memory malware detection in resource-constrained IoT devices for smart city applications. *Sensors*. https://doi.org/10.3390/s23115348

[30] Shaukat, K., Luo, S., Chen, S., Liu, D. (2020). Cyber threat detection using machine learning techniques: a performance evaluation perspective. 1st Annual international Conference on Cyber Warfare and Security. ICCWS 2020 Proceedings. https://doi.org/10.1109/ICCWS48432.2020.9292388

[31] Shroff, J., Walambe, R., Singh, S.K., Kotecha, K. (2022). Enhanced security against volumetric DDoS attacks using adversarial machine learning. *Wireless Communications and Mobile Computing*. https://doi.org/10.1155/2022/5757164

[32] Srivani, M., Abirami, M., & Mala, T. (2023) Cognitive computing technological trends and future research directions in healthcare–A systematic literature review. *Artificial Intelligence in Medicine* 138, 102513. https://doi.org/10.1016/j.artmed.2023.102513.

[33] Tan, Z., Beuran, R., Hasegawa, S., Jiang, W., Zhao, M., & Tan, Y. (2020). Adaptive security awareness training using linked open data datasets. *Educ. Inf. Technol*., 25(6), 5235–5259.

[34] Torres, J.M., Comesana, C.I., & García-Nieto, P.J. (2019). Machine learning techniques applied to cybersecurity. *Int. J. Mach. Learn. Cybern*, 10(10), 2823–2836.

[35] Truong, T.C., Zelinka, I., Plucar, J., Candík, M., & Sulc, V. (2020). Artificial intelligence and cybersecurity: past, presence, and future. In: Artificial intelligence and evolutionary computations in engineering systems. pp. 351-363.

[36] Vanjire, S. & Lakshmi, M. (2021). A behavior-based malware detection system approach for mobile security using machine learning. In Proceedings of the 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India. September 24-26, 2021. pp. 1-4.

[37] Wazid, M., Das, A.K., Chamola, V., Park, Y. (2022). Uniting cyber security and machine learning: Advantages, challenges, and future research. In *ICT express* (pp. 313–321). Korean Institute of Communication Sciences. https://doi.org/10.1016/j.icte.2022.04.007.

[38] Welukar, J.N., & Bajoria, G.P. (2021). Artificial Intelligence in Cybersecurity - A Review. *International Journal of Scientific Research in Science and Technology*, 8(6), 488-491. https://doi.org/10.32628/IJSRST218675

[39] Wilkins, J. (2018). Is artificial intelligence a help or hindrance? *Network Security,* (5), 18-19.

[40] Zhang, Z., Ning, H., Shi, F., Farha, F., Xu, Y., Xu, J., Zhang, F., & Choo, K.K.R. (2022). Artificial intelligence in cyber security: Research advances, challenges, and opportunities. *Artif. Intell. Rev*., 55, 1029-1053.