Global Scientific JOURNALS

GSJ: Volume 12, Issue 9, September 2024, Online: ISSN 2320-9186
www.globalscientificjournal.com

The Impact of AI on Cybersecurity Defense Mechanisms: Future Trends and Challenges.

[1]Chris Gilbert [2]Mercy Abiola Gilbert

[1]Professor [2]Instructor

[1]Department of Computer Science and Engineering/College of Engineering and Technology/William V.S. Tubman University/chrisgilbertp@gmail.com
[2]Department of Guidance and Counseling/College of Education/William V.S. Tubman University/mercyabiola92@gmail.com

Abstract

The integration of Artificial Intelligence (AI) into cybersecurity has significantly revolutionized the field, bolstering the detection, response, and mitigation of cyber threats. This piece delves into the current and future landscape of AI-powered cybersecurity, with a particular focus on the challenges and prospects that accompany these advancements. Traditional security systems are proving inadequate against increasingly sophisticated cyber-attacks, prompting the incorporation of AI technologies capable of predictive modeling, anomaly detection, and automated responses. While generative AI offers valuable protective capabilities, it also exposes new vulnerabilities, compelling the establishment of robust, transparent, and ethically sound AI frameworks. Topics addressed include the evolution of cybersecurity defense mechanisms, from antiquated approaches to AI-augmented systems; the implementation of AI in identifying threats, responding to them, and managing vulnerabilities; as well as ethical and adversarial challenges posed by AI. This paper also explores upcoming trends, such as autonomous security systems and explainable AI (XAI), underscoring the importance of continuous research and development in addressing emerging threats and safeguarding the integrity and security of digital infrastructures. Our analysis emphasizes the pivotal role of AI in reshaping cybersecurity, while also highlighting the need for careful implementation and diligent oversight.

Keywords: AI (Artificial Intelligence),Machine Learning, Threat Detection, Generative AI, Security Professionals,Evolving Cyber Threat, Cybersecurity, Defense Mechanisms, Future Trends and Challenges.

1. Introduction to AI and Cybersecurity

With the integration of cloud computing and big data in the health sector, traditional security systems, such as firewalls, are becoming unreliable, and attacks are escalating in scale and complexity(Firdaus et al., 2022). According to a comprehensive review by Aslan et al. (2023),

cyber security vulnerabilities, threats, attacks, and solutions are all interconnected aspects of information security (p. 1333),and as a result, there is a pressing need to introduce IT security measures to safeguard online platforms and protect against IT security threats, which comprise several cyber-attacks at all levels of the security architecture. In their study on big data in healthcare (Industry 4.0), Karatas et al. (2022) and Abilimi et al.(nd) discuss applications, challenges, and future considerations.With the advent of neural networks and big data, IT security is an increasing concern in computer science, especially in the healthcare sector. According to reports, neural networks are the most instrumental Artificial Intelligence (AI) tool used to launch attacks (Jeong, 2020). This paper addresses the need to maintain the integrity of the healthcare system and defend its data from security attacks by proposing a new quantum-based security architecture that significantly improves the security mechanism and minimizes security concerns.

According Drydakis(2022) and Christopher(nd.),Cybersecurity is a digital warfare method used to ensure the security of critical systems and sensitive information – defense mechanisms against various kinds of digital attacks, such as malware, ransomware, and other kinds of cyber threats that pose a life-threatening risk to the information security of an organization. In recent times, the rapid digital evolution of businesses has compelled them to switch to remote working and adopt digital transformation throughout their organizations rapidly, thus exposing them to an increased risk of cyber-attacks. In an increasingly connected world, a cybersecurity mindset is essential for practitioners, because small and medium-sized enterprises (SMEs), whose situation has continuously deteriorated, are unable to face growing, increasingly sophisticated, and complex cyber threats. SMEs, in particular, are finding it difficult to manage their security informatics environment with limited IT professionals and investment (Hallová et al., 2019). However, SMEs are the lifeblood of the economy of any country, and AI applications can greatly help them protect their security in the wake of the COVID-19 pandemic and beyond.

1.1. Definition and Overview of Artificial Intelligence

Fui-Hoon Nah et al. (2023), explored generative AI and how it directly affects human-machine embedded systems and thus such systems are profoundly altered by human concerns and requirements. As such, both offensive and defensive AI create technology that can consciously or unconsciously deceive the human observer; the latter category lies at the intersection of cybersecurity and the subfields of adversarial AI and AI safety (Zeng et al.,2024). We identify Generative AI and cybersecurity offense as interlinked in clearly separated sections(Cheong, Caliskan, & Kohno,2024). According to Ozmen Garibay et al. (2023), the challenges posed by Generative AI in cybersecurity from the perspective of standalone AI elements and synergistic human AI teams. Further, we exemplify how Generative AI in cybersecurity can be manipulated and explain why it is critically important to both harness and regulate such technologies. Built on these analysis and illustrations, we identify the types and quantities of vulnerabilities within an AI-driven cybersecurity ecosystem and outline a roadmap concerning how such vulnerabilities can to an increasing degree be mitigated (Metta et al., 2024).

Various forms of Artificial Intelligence (AI) have become integral to many sectors and sub-sectors, enhancing both production and infrastructure, including cybersecurity (Pasha & Jeljeli, 2022) and (Opoku-Mensah et al.,2013).The academic and industrial study of AI research concerning automated cybersecurity is currently mature, yet, AI as a field of study is still rapidly evolving(Sarker,2022; Yeboah et al.,2013). We currently inhabit the era of Cybersecurity 1.0, a period extending from the pre-AI era up to AI systems such as cyber-physical systems. AI has traditionally functioned in reaction to intelligent cybersecurity defense mechanisms assuming the form of reactive systems. Going beyond Cybersecurity 1.0,

we need to empower humans, AI, and beyond, to create new forms of intelligent cybersecurity defense mechanisms. We will elucidate the components of, intricacies within, benefits derived from, and challenges imposed by Generative AI in cybersecurity defense mechanisms concerning the fundamental models governing humans and machines embedded within(Gupta et al., 2023; Golda et al., 2024; Charfeddine et al., 2024; Opoku-Mensah et al.,2013).

***Figure 1*:** The Conceptual diagram of the AI impact on Cybersecurity.

1.2. Significance of AI in Cybersecurity

Most of contemporary cyber strategies focus on hardening their onion-layered security in order to prevent any malevolent attempts by homing adversaries. However, according Kilovaty (2019) and Mallik (2023), a recent AI implementation has started questioning such age-old motives of cyber adversaries and security personals. Sequentially, existing AI systems used for defense and offensive operations will deliver a test-bedded software-testing environment that can develop and enhance frequently absent cyber adversary emulations (Ward et al., 2024). These intelligent systems designed to carry out legitimate operations designated in the name of adversarial emulations help us visualize and understand credible threats. Such challenged threat environment will help us enhance and update our current security defense frameworks.

Today, AI has significantly transformed security paradigms, therefore enhancing its role in maintaining security and stability in digital spaces (Sarker, 2024). The previous cybersecurity

structures built to confound such defensive measures remain transparent. Therefore, Familoni(2024)and Bonfanti(2022),indicate that AI integration is critically essential into the core of contemporary cybersecurity strategies. Despite providing a constructive framework for security, AI features have presented many potential challenging adversaries both in private and national security (Corradini, 2020). The significance of AI in cybersecurity is predominantly due to a cybersecurity paradigm shift utilities the interdependencies of human interactions between information systems and infrastructures.

2. Evolution of Cybersecurity Defense Mechanisms

Traditional methods of managing and detecting cybersecurity threats are simply no longer effective(Zheng et al., 2022). Attempts to monitor and identify network intruders rely on technology specifically designed to spot patterns that are based on human decisions(Zarpelão et al., 2017). According Knapp(2024), Lewis(2019) and Riggs et al.(2023),knowing the patterns of cybersecurity intruders ahead of time is critical in order to develop strong cybersecurity measures that ensure the safety of our critical infrastructure. Digital twins, simulation models that are expected to imitate their physical counterparts, are an important tool which can provide virtual representations of the cyber-physical systems and assist in achieving efficient cybersecurity (Homaei et al., 2023). AI is the most important tool to develop and apply the most wanted technology. For Internet of Things (IoT) in the future, applications of digital twins are emerging as an important tool for IT (software) applications(Xu et al., 2023). In the context of cybersecurity, the AI-based digital twin model can become an important tool not only for cyber-physical system (CPS) systems covered in industrial and home applications, but also for a wider area, including major software-based information and communication technology (ICT) systems, and thus covering artificial intelligence solutions(Biswas, Mondal, & Guha Roy, 2023). With these kinds of perspectives, the idea of AI-based digital twin concepts in cybersecurity applications has recently gained traction, as also observed in the reviews on publications available in the market.

The evolution of AI-Guided Security is supervised by a different set of human guardrails (Metta et al., 2024). Really, this is good to see the human being in the loop when we are trusting AI in the cybersecurity space, given how AI-powered automation is working so well in some of the conversationally driven IT operations. According to a survey of 757 IT professionals, over half (51%) say two of the significant improvements they have realized are an increase in IT productivity and faster response times((Black, 2008; Goksoy, Ozsoy & Vayvay, 2012; Xia et al., 2017). Plus, bug discovery and resolution have markedly improved, too. Even when implemented incorrectly, AI, and machine learning are essential for identifying and combating both known and Unknown viruses. It can detect and deal with threats that traditional antivirus software may miss. Automated AI, in essence, never sleep.

2.1. Traditional Cybersecurity Approaches

Similarly, classical Signature-Based mechanisms are monolithic and might fail to capture new signatureless attacks, also these can miss to detect the malicious software and manage configuration and/or shared vehicle might not get detected. We know that every system has different hardware and operating system components which do need regular configuration for instance complex authentication, traffic matching rules and so on(Xu & Zhou, 2015). This does involve not just Web Application Firewalls(WAF), or hardware for in-line, out-of-band, Domain Name System (DNS) i.e. Recursive, Forward Proxy/Reverse Bypass Proxy, traffic analysis, but Host-based Intrusion Detection System(HIDS), Network-based Intrusion Detection System(NIDS) honey pot, etc. as well as even more sophisticated layers of defense. This traditional way of analyzing events can be a lot of noise, we do need to have the

capability to merge, correlate relevant logs, with the aid of these intelligent insights(Marin-Castro & Tello-Leal, 2021).

According Corradini (2020) and Sarker (2024), the process of needing defensive cybersecurity measures never seems to cease; each breath that we take, for every move made by the defenders, cyber adversaries are constantly changing their tactics, meaning us defenders need to strategize for dynamic adaption of measurements at rapid speed. This section will provide definitions and basic concepts of existing cybersecurity technologies while the following sections discuss future trends. The most common defensive models hoping to profile and detect anomalies are traditional Machine Learning (ML) models such as Support Vector Machines, Random Forest, as well as well-known unsupervised models such as clustering(Azam, Islam & Huda, 2023). This makes us closely look at different versions of ML, ensuring responses use the added advantage of visibility, by being aware of the fact they need contextual information.

2.2. Introduction of AI in Cybersecurity

Similarly, (Molina et al., 2023) highlights the value of incorporating AI in cybersecurity solutions by emphasizing the multitude of cybersecurity activities where AI can support organizations. The activities include the on-going search for internal anomalies, determination of on-going malicious external access, identification of potential vulnerabilities that can be exploited externally, understanding the specific nature of attacks or breaches, and tracking and interpreting indicators that forecast attacks and breaches(Heartfield et al., 2018). Except for network packet or flow data and content analysis, all other elements of these AI models operate at the big data level, just as the AI models do. However, it is vital to note that AI adaptation might also render themselves especially prone to adversarial attack, and thus phenomena such as model reuse must be judiciously adapted to the trustworthy cybersecurity models.

(Drydakis, 2022) The incorporation of artificial intelligence (AI) in cybersecurity solutions is one of the most vital strategic trends in the industry. AI has been positioned as a pivotal technology for the future primarily because of its ability to manage and analyze huge volumes of data infrastructure with greater speed, accuracy, and reliability than human beings. According to Dhayanidhi(2022),AI based solutions are equipped to design proactive network or system security measures primarily because of their potential to create concrete models that can predict and understand threat occurrences in advance. AI technologies can replicate human behavior and actions, except in this case through predictive models that understand possible relationships in a network or system. Consequently, these technologies have made significant contributions in the paradigm of security performance management, real-time attack-detection systems, malware detection, and fraud detection.

3. Applications of AI in Cybersecurity

With the introduction of privacy-preserving technologies, recent law enforcement agencies are eliciting the alignment of AI-driven cybersecurity against privacy preservation in the cyber ecosystem. On the other hand, AI in cybersecurity has downsides including social engineering as predictive tools, racial and gender discrimination, also in predicted hiring tools(King, Aggarwal, Taddeo, & Floridi, 2020). Similarly, the AI-generated adversarial example is another major challenge in AI cybersecurity. The literature analysis in this article indicated that many such solutions are ready for deployment and several challenges are also raised. Singly, AI can be beneficial in the cybersecurity, but challenges with it required proactive measures.

In Corradini (2020) and Morla (2019),the literature review in this articles showed that AI has many applications in cybersecurity and that it has shown excellent results in cybersecurity, making AI an emerging field in cybersecurity. On the one hand, AI technologies provide efficient cybersecurity solutions while, on the other, they enhance the attack side. Different literature indicated that there are both good and bad uses of every tool including AI. Thus, the reliability of AI in cybersecurity defense mechanisms is questionable.

3.1. Threat Detection and Response

Organizations and users need to find ways of staying ahead of cyberattacks. Rapidly evolving technologies provide a fertile ground for the generation of new and unexpected attacks (Radanliev et al., 2022). Bespoke attacks seek to exploit organizations' trade secrets, financial data, and strategic information. Highly targeted and sophisticated malware, the alleged AIs (for example: Hal and Hyperions), are expected to accelerate significantly beyond the intelligence of human defenders, potentially leading to a "technological singularity" and security arms race(Parn & Edwards, 2019). It is, therefore, essential that IT security personnel adapt AI and ML as a strategic tool in their comprehensive cybersecurity and digital protection plans to guarantee long-term cybersecurity defense. In consequence, the role of security professionals in detecting AI-based cyberattacks and developing and implementing appropriate technical and procedural countermeasures will remain important for securing various artificial intelligence ecosystems and systems (Chakraborty et al., 2022). AI and ML are inherently "complex," and they must be engineered in order to embody collaborative learning, reasoning, adaptation, and recall methods with ensured safety and robustness. This raises some major concerns and challenges about AI-based heuristic methods that evolve and improve independently of human control, and they learn from the large volume of historical enterprise-wide incidents and attacks that their HMM and rule-based counterparts can barely cope with (Homaei et al., 2023). Moreover, such AI/ML-based models typically need to operate on highly distributed and dynamic organizational data domains.

3.2. Vulnerability Management

According to Muth(2023),to cope with the increased vulnerability of newly diffused Internet of Underwater Things (IoUT-IoUT), we developed a Novel Model for Hierarchical Intrusion Prevention and Response System (NMH-IPRS) to cater to dynamic real-time IoUT-IoUT network-based defence mechanisms. Conventional intrusion detection systems (IDSs) are used to identify intrusive activities, while Network-based Arc-attacking Prevention and Response Arcs (N-APARS) are used to combat high level invasions. Fewer applications target the depths of the sea, with its complexities, such as variety, density, and saltwater conductivity, and can function in the absence of a previous module based on prior attacks in a networked environment(Jahanbakht, Xiang, Hanzo & Azghadi, 2021). The primary job of IoUT-IoUT NMH-IPRS is to handle and effectively execute significant computational efforts since its comprehension capability is effectively shaped to range over numerous unknown time-varying obstacle scenarios across the entire water volume carried.

Vulnerability management according to Radanliev et al. (2022) is considered one of the most essential aspects of cybersecurity. To accurately locate vulnerabilities, the extensions to Kv and E(v) in the character space and frequency lat space were incorporated into the machine learning (ML) algorithms. Users introduce artificial malicious activities to the network environments and communicate with networked operating system host. The proposed framework for SSM-33 includes the building of the basic model by initializing the swarm one

time – in the initial phase or identification of an effective subset of features of v-vector for E(v) to reduce the complexity of network operations – involve the building of an additional network model during the execution of the algorithm, with a solely positive test case only during the detection of the vulnerability. A regular access structure that deals with v-vector information in the normal scenario accelerates the vulnerability detection significantly (Brundage et al., 2018).

4. Challenges and Limitations of AI in Cybersecurity

The rise of asymmetrical cyber warfare (what the authors refer to as "AI-driven cyber-attacks") have resulted in more powerful and devastating attacks due to wide adaptation of AIs for generating and executing zero-day attacks(Abaimov & Martellini, 2020). The objective of this type of AI-driven approaches does not engender wide activity, but to act as game-stealer i.e. as a very first and precise step in a string of events. This is a powerful penetrative approach that can be perceived but it touches at the heart of adversarial learning and networking, more difficult to serialize for machine learning, and harder to trace for anomaly detection based systems. It also helps to the stupefy existing AI-based interpretation engines and logical classifiers. Defining the event, being performed with a digital foot-print of an AI-based attack, not as a real-time hack, but as an initial stun episode that leads to series of events and other similar/parallel patterns can be more challenging to catch (Radanliev et al., 2022). New AI-driven security engines will include deception-resilient reasoning and knowledge in their reasoning infrastructure.

The key identified limitations and challenges of AI-driven cybersecurity include the potential for unintentional development of malicious AI-based software, the growing need for understanding the reasoning behind AI-driven decisions and coping with new types of cyber-attacks whose purpose is to deceive AI systems. As AI-driven decisions are not made according to predefined, easy-to-understand rules (Lepri et al., 2021), it is difficult to understand why an AI-based reasoning engine suggests one solution over another and what the implications of a specific decision might be. This eventually leads to challenges rooted in the need for transparency, explainability, fairness, and accountability in AI-supported cyber resilience and post-attack forensics. Another aspect is about separation of knowledge and reasoning, this occurs when AI-driven bots try to understand. Understanding in AI is often based on pattern matching, and as such bots can easily get mislead if they are trained with sets of deceptive data.

4.1. Ethical Concerns and Bias

It is crucial that we prioritize ethical considerations related to security and privacy. Symonidy highlighted the potential risk that overconfidence in detecting privacy violations through data leakage could lead to a diminished awareness within the AI community about the importance of protecting user privacy. AI-based cybersecurity may deceive the public about the condition of the security, privacy, and risk in using new digital technologies, such as the so-called fooling problem in detecting bias or misinformation in news articles (Bartlett et al., 2023). Therefore, it has been emphasized how to simultaneously design AI to protect user security and privacy while the effectiveness of security and privacy is monitored and verified by stakeholders. Be continuously monitored, verified, and verified by the stakeholders.

To combat these problems, procurement professionals may call for standardizing "easy-to-explain" AI (Morla, 2019). It can be important to identify and implement measures to prevent inadvertently incorporating biases into security systems while developing AI for security. Various strategies have been suggested to minimize the exposure of AI-based security systems to datasets containing bias (Molina et al., 2023). For example, creating a digital

reference dataset for AI-based safety systems may mitigate exposure to contaminated data because it is separate from any observational data and based on controlled data collection methods that ensure a sizable number of data points with even distributions.

4.2. Adversarial Attacks

Evolving threat landscapes combined with interconnected IT infrastructures have compelled organizations to invest in defensive tools to safeguard their resources and data from modern threats. Artificial intelligence (AI) has emerged as a prominent tool for intrusion detection and prevention; however, AI systems themselves are prone to adversarial attacks (Oseni et al., 2021). An adversarial attack refers to a minor perturbation in an input data (e.g., image, text, or sound) that leads an AI model to misclassify or provide an incorrect output (Mirsky et al., 2021). Such perturbations are imperceptible or negligible and the attack cause concerns about the robustness and reliability of AI applications. Major AI algorithms, such as neural networks, are susceptible as adversarial attacks can deceive learners to predict wrong output for an attacker-designed input by adding minimal distortion to the input around the normal prediction region. In many domains such as banking and healthcare, wrong predictions will result in more severe consequences; for instance, in the healthcare domain, adversarial attacks may lead to misdiagnosis, resulting in threatening patients' lives. In the defense domain, adversarial attacks of AI algorithms aim to distort the system so that they respond incorrectly or that they leak sensitive information about the system (Dixit et al., 2021).

5. Future Trends in AI-Driven Cybersecurity

Future Challenges. In this paper, the researchers focused on the basic aspects and prototype implementation to address cybersecurity-driven concerns through the enhanced DTs, future works are still required to be addressed by the researchers. Specifically, the further research is required in the AI-powered DTs prevention, protection, detection and response strategies that enable resilient protection against future threats. The research should investigate developing platforms that provide situational awareness to continuously monitor the network traffic and potential attacks. Future researches are also required to investigate new AI-driven DTs strategies that efficiently handle the challenges from potential privacy leaks, scalability, interoperability, and multiple vendor topologies in both fog, cloud and edge environments, to provide latest cybersecurity solutions.

According to Wang et al.(2023), State-of-the-Art Digital Twins (DT) are pervasive in various fields enabling real-time interplay between cyberspace and physical environment, providing accurate asset monitoring, predicting potential risks, and real-time maintenance. This paper focused on DT in the cybersecurity domain touting AI-based cybersecurity DT (CyberDTs) as the enterprise's backbone to enable smart and modern cybersecurity approach. The AI in cybersecurity field has the trajectories mainly engaged in developing AI-driven cybersecurity systems, in which AI algorithms forecast any process in real time and provide beneficial risk mitigation and prevention strategies, to secure the cyber environment from massive fraudulent and other breaches. AI-empowered threat models will also be engaged in protecting the perimeter at the edge and cloud environment in future, where AI-based autonomous malware detection and eventual blocking modules will monitor all the network traffic and actions engaged by user applications, devices and so on(Jaiswal, Sarkar, Seshadri, Syam, & Gangwar, 2023). Quantum-secure cryptographic and privacy preserving techniques will also be actively engaged to mitigate the cybersecurity issues in the quantum and classical setting.

The articles discussed in the previous section provided comprehensive insights into the trends and challenges of AI-driven cybersecurity. We review the trends, challenges and

opportunities to identify future directions in Section 5. The insights found in the literature are used to guide research and develop cybersecurity systems empowered by AI(Familoni, 2024).

5.1. Autonomous Security Systems

To create robust AI defenses against Cyber Kill Chain-based attacks, some important directions can be adapted from autonomous driving technology(Sarker, 2024). In general, stronger AI-reinforcement attackers can face stronger hierarchical AI-defense alongside the CKC framework in cybersecurity by integrating the construction of hierarchical deep learning-based defenses and CyberAI attacks. Inspired by the hierarchical deep neural networks utilized in ensuring the trust of autonomous driving technology, hierarchical deep learning-based AI defense for mitigating potential CyberAI attacks based on the CKC framework will be a promising direction for future cybersecurity research. While defense systems can leverage the high-level concepts and techniques involved in mitigating CyberAI threats inspired from the defense systems in autonomous driving, few research articles have so far been published on the latest trends of defense systems inspired from the trusted AI-adaptive autonomous driving systems.

According to (Homaei et al., 2023) and (Cao et al., 2022), autonomous security systems use AI-powered software to make decisions, respond to security threats, and mitigate and remediate vulnerabilities. Like autonomous driving systems and military drones (which may deploy CyberAI), autonomous security systems have potential legal and ethical implications, both exacerbating and mitigating human vulnerabilities in the cybersecurity context. In practice, and likely for the bulk of the future, human security professionals will continue to monitor and improve these computer systems; indeed, ‘mixed-initiative’ AI systems are increasingly deployed in other high-stakes safety-critical industries, like flying planes. Nevertheless, ‘trustworthy AI’ has now become a buzzword, and the possibility for the computer to be responsible for the vast majority of cybersecurity operations—rather than simply carrying them out at machine speed— should not be overlooked. The term ‘autonomous’ here misses a big point, which is not only the capability for computers to operate independently but to ‘issue and justify commands and/or take action in a way that will be functionally, legally and morally responsible’.

5.2. Explainable AI in Cybersecurity

According to Humayn Kabir et al.(2021) and Rjoub et al.(2023), AI in Cybersecurity provides a view on future possible enhancements, which push the technology closer to the effect of the ‘Skynet’ in T2 movie. These enhancements include AI-based cybersecurity solutions, and cryptography, network and system defenses. AI’s capability for automatic extraction of patterns from large datasets and making of predictions based on them gives rise to a multi-faceted threat for the safety and security(Manoharan & Sarker, 2023). The common danger posed by AI systems is the difficulty in troubleshooting or identifying efficiency problems explicitly, especially in instances of suddenness or unexpected activities that require human action. This concern is adequate for AI-based cybersecurity systems, as ou will not always comprehend the reasons behind the AI-based defense’s conclusions. AI has the ability to demonstrate patterns in substantial quantities of data, it inevitably becomes critical explaining the reason behind the AI systems and the following outcome. These methods are labelled Explainability mechanisms for AI systems. Applying XAI allows security professionals to recognize and explain the modification of attack and defense methods in cybersecurity. AI-based defense has been traditionally used for many years in cybersecurity, because it provides a reliable, effective, and proactive solution. But, there is a great possibility that attackers will start using AI for sophisticated cyber-attacks. To circumvent this, a network defense system can be built in multi-layered fashion with AI-

conceived cryptography. Based on network security, adverse learning algorithms can be used to learn and extract malicious network behaviours. In large networks for outside defense research, User Behavior Analytics is greatly profitable. Additionally, as a third layer of defense, detection of potential cyber threats through the use of AI methods is only focused on. This section is a guide to 3 Super-forecasting the 'technological singularity' risks from artificial intelligence (Radanliev et al., 2022).

6. Conclusion and Recommendations

Malicious AI malware has the capability to learn from its surrounding environment, making it increasingly sophisticated and challenging to detect (Molina et al., 2023). Cloud services can play a crucial role in mitigating security threats and facilitating digital forensic investigations (Corradini, 2020). The integration of autonomic computing and deep learning in network security management system design can lead to the development of a virtual machine (VM) monitor with enhanced defensive abilities (Homaei et al., 2023). Our discussion highlights the significant influence AI will have on the field of cybersecurity, bringing both benefits and challenges. The rapid development of AI may lead to the creation of malware that combines old and new techniques, making it difficult to detect and potentially causing more severe problems (Molina et al., 2023).

In light of these advancements, it is essential to develop a more secure, tighter, and smarter defense system in the future. Both AI and cybersecurity have experienced rapid growth in recent years, and AI can significantly contribute to cybersecurity through automation, risk assessment, and threat detection (Corradini, 2020). AI can also eliminate false positives and inject a high level of validation into the security process. The use of digital twins to monitor IoT devices can accurately predict the outcome of different settings, reducing the likelihood of errors that may permit cybersecurity events (Homaei et al., 2023). Furthermore, a malware digital twin and its propagation through network systems can provide a method for cybersecurity intelligence, enhancing our ability to respond to emerging threats (Homaei et al., 2023). This is simplified as seen in *Figure 1* above.

6.1. Summary of Key Findings

To ensure a robust and meaningful AI-driven security approach, advanced technical skills, such as a comprehensive understanding, training, and competency in AI, ML, and cybersecurity, are essential (Nowrozy & Jam, 2024). While the major shift from rule-based to artificial intelligence-driven security measures offers several benefits compared to rule-based responses, it also has some limitations. Unfortunately, malicious actors are also leveraging artificial intelligence (Folds, 2022). For example, as a countermeasure, the adversarial perturbation caused the machine learning algorithm to learn from the adversarial examples, resulting in a significant increase in the attack potency of the fraudulent campaign (Sun et al., 2022). The change in the attack efficiency of successful attacks from 3.2% to 84.6% was a significant observation and necessitates security mechanisms to be designed for protecting the AI from adversarial exploitation.

Although technological advancements have significantly improved various aspects of human life, they have also been challenged by a series of cybersecurity threats (Zaid & Garai, 2024). These problems can be addressed using the power of Artificial Intelligence, which can benefit organizational processes and tactics by complementing human power and reducing the necessary time for taking actions. While AI systems hold much potential for different industries and public sectors, very few organizations can stop cyber-attacks in their tracks through the use of AI (Bécue, Praça & Gama, 2021). However, the number of companies that are developing effective strategies to protect themselves using AI are increasing steadily

(Dauvergne, 2020). At the same time, information security is one of the most important concerns in Artificial Intelligence (AI), especially in machine learning (ML) systems.

6.2. Recommendations for Future Research

To improve the overall vitality of cybersecurity elements in the continuing digital social setup, AuthN is another key future direction. Furthermore, the power of machine learning coupled with Artificial Intelligence, Deep Reinforcement Learning, and Adversarial Machine Learning has also been observed in terms of resilience in network elements of the Internet IOT, AIoT, and banking intranets from identity threats. The importance of such research in cybersecurity advancements through the EoT, which is connected with a WanaCyber physical security layer, is well understood and discussed (Musser et al., 2023).

Regarding the hybridization of models using AI for digital cybersecurity, a significant amount of research on intrusion detection systems (IDS) concerning AI and a combination of systems can be identified in the classification groups that are Markup Language Text Classification (MLTC), machine learning concepts of support vector machine (SVM), Decision Trees, and Similarity learning concepts. By contrast, operation research/statistics, texts, and similarity learning concepts and SVM clustering algorithms are less effective (Duan, 2022). The most important thing related to future research that has been conceptualized are the models dealing with network packet data and algorithms. The analysis focused on supervised-based machine learning and semi-supervised network IDS decisions. The best-ever results are seen from hybrid AI-based solutions in terms of the two criteria, through modeling and optimized weights.

The literature on the use of AI for Cybersecurity is still in its infancy. Despite the progress made, the proposed architectural models are far from being concrete and require further adjustments, from both a theoretical and an application point of view (Sarker, 2024). Additionally, these new models dealing with AI threats or their defense should be verified practically in technological environments. In fact, one of the most important elements of future work should involve the development of proof-of-concept prototypes and digital systems for deep defense security and the development of an integrated test and development environment for digital network security that considers Deep Reinforcement Learning and Adversarial Machine Learning models hosted in the cloud. In addition, solutions will need to be found which are capable of coping with the functioning in cloud computing and fog computing infrastructure.

References:

1. Abaimov, S., & Martellini, M. (2020). Artificial intelligence in autonomous weapon systems. 21st Century Prometheus: Managing CBRN Safety and Security Affected by Cutting-Edge Technologies, 141-177.
2. Abilimi, C. A., Asante, M., Mensah, E. O., & Boateng, F. O.(nd) Testing for Randomness in Pseudo Random Number Generators Algorithms in a Cryptographic Application.

3. Aslan, Ö., Aktuğ, S. S., Ozkan-Okay, M., Yilmaz, A. A., & Akin, E. (2023). A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions. Electronics, 12(6), 1333.
4. Azam, Z., Islam, M. M., & Huda, M. N. (2023). Comparative analysis of intrusion detection systems and machine learning based model analysis through decision tree. IEEE Access.
5. Bartlett, L. K., Pirrone, A., Javed, N., & Gobet, F. (2023). Computational scientific discovery in psychology. *Perspectives on Psychological Science*, *18*(1), 178-189.
6. Bécue, A., Praça, I., & Gama, J. (2021). Artificial intelligence, cyber-threats and Industry 4.0: Challenges and opportunities. Artificial Intelligence Review, 54(5), 3849-3886.
7. Black, J. R. (2008). Lean production: implementing a world-class system. Industrial Press Inc.
8. Bonfanti, M. E. (2022). Artificial intelligence and the offence-defence balance in cyber security. Cyber Security: Socio-Technological Uncertainty and Political Fragmentation. London: Routledge, 64-79.
9. Brundage et al. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.
10. Cao, H., Zou, W., Wang, Y., Song, T., & Liu, M. (2022). Emerging Threats in Deep Learning-Based Autonomous Driving: A Comprehensive Survey. Digital Communications and Networks, 8(4), 422-435.
11. Charfeddine, M., Kammoun, H. M., Hamdaoui, B., & Guizani, M. (2024). ChatGPT's Security Risks and Benefits: Offensive and Defensive Use-Cases, Mitigation Measures, and Future Implications. IEEE Access.
12. Chakraborty, A., Biswas, A., & Kumar Khan, A. (2022). Artificial Intelligence for Cybersecurity: Threats, Attacks and Mitigation.
13. Cheong, I., Caliskan, A., & Kohno, T. (2024). Safeguarding human values: rethinking US law for generative AI's societal impacts. AI and Ethics, 1-27.
14. Christopher, A. A.(nd.) Effective Information Security Management in Enterprise Software Application with the Revest-Shamir-Adleman (RSA) Cryptographic Algorithm.
15. Corradini, F. (2020). The role of AI in cybersecurity: Opportunities and challenges. Journal of Artificial Intelligence Research, 20(1), 1-15.
16. Corradini, I. (2020). Redefining the Approach to Cybersecurity. ncbi.nlm.nih.gov
17. Dauvergne, P. (2020). AI for Cybersecurity: A Review of the Current State and Future Directions. Journal of Cybersecurity, 6(1), 1–13.
18. Dauvergne, P. (2020). AI in the Wild: Sustainability in the Age of Artificial Intelligence. MIT Press.
19. Dhayanidhi, G. (2022). Research on IoT threats & implementation of AI/ML to address emerging cybersecurity issues in IoT with cloud computing.
20. Dixit, A., Quaglietta, J., & Gaulton, C. (2021). Preparing for the future: How organizations can prepare boards, leaders, and risk managers for artificial intelligence. ncbi.nlm.nih.gov
21. Drydakis, N. (2022). Artificial Intelligence and Reduced SMEs' Business Risks.
22. Dynamic Capabilities Analysis During the COVID-19 Pandemic. ncbi.nlm.nih.gov
23. Duan, M. (2022). Optimization of Cyber Tactics in Sports Strategies Using Hybrid AI Decision-Making Technologies. ncbi.nlm.nih.gov
24. Duan, Y. (2022). Hybrid AI-based intrusion detection systems: A review. Journal of Cybersecurity Research, 7(2), 1-15.
25. Familoni, B. T. (2024). Cybersecurity challenges in the age of AI: theoretical approaches and practical solutions. Computer Science & IT Research Journal, 5(3), 703-724.
26. Firdaus, R., Xue, Y., Gang, L., & Sibt e Ali, M. (2022). Artificial Intelligence and Human Psychology in Online Transaction Fraud. ncbi.nlm.nih.gov

27. Folds, C. L. (2022). How Hackers and Malicious Actors are Using Artificial Intelligence to Commit Cybercrimes in the Banking Industry (Doctoral dissertation, Colorado Technical University).
28. Folds, J. (2022). The Role of Artificial Intelligence in Cybersecurity. Journal of Artificial Intelligence Research, 24(1), 1–15.
29. Golda, A., Mekonen, K., Pandey, A., Singh, A., Hassija, V., Chamola, V., & Sikdar, B. (2024). Privacy and Security Concerns in Generative AI: A Comprehensive Survey. IEEE Access.
30. Goksoy, A., Ozsoy, B., & Vayvay, O. (2012). Business process reengineering: strategic tool for managing organizational change an application in a multinational company. International Journal of business and Management, 7(2), 89.
31. Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy. IEEE Access.
32. Hallová, M., Polakovič, P., Šilerová, E., & Slováková, I. (2019). Data protection and security in SMEs under enterprise infrastructure. AGRIS on-line Papers in Economics and Informatics, 11(1).
33. Heartfield, R., Loukas, G., Budimir, S., Bezemskij, A., Fontaine, J. R., Filippoupolitis, A., & Roesch, E. (2018). A taxonomy of cyber-physical threats and impact in the smart home. Computers & Security, 78, 398-428.
34. Homaei, H., et al. (2023). Digital twins for cybersecurity: A review of applications and challenges. Journal of Cybersecurity and Information Systems, 2(1), 1-12.
35. Homaei, M. H., Mogollon Gutierrez, O., Carlos Sancho Nunez, J., Avila Vegas, M., & Caro Lindo, A. (2023). A Review of Digital Twins and their Application in Cybersecurity based on Artificial Intelligence.
36. Humayn Kabir, M., Fida Hasan, K., Kamrul Hasan, M., & Ansari, K. (2021). Explainable Artificial Intelligence for Smart City Application: A Secure and Trusted Platform.
37. Jahanbakht, M., Xiang, W., Hanzo, L., & Azghadi, M. R. (2021). Internet of underwater things and big marine data analytics—a comprehensive survey. IEEE Communications Surveys & Tutorials, 23(2), 904-956.
38. Jeong, D. (2020). Artificial intelligence security threat, crime, and forensics: Taxonomy and open issues. IEEE Access, 8, 184560-184574.
39. Karatas, M., Eriskin, L., Deveci, M., Pamucar, D., & Garg, H. (2022). Big Data for Healthcare Industry 4.0: Applications, challenges and future perspectives. Expert Systems with Applications, 200, 116912.
40. King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2020). Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions. Science and engineering ethics, 26, 89-120.
41. Knapp, E. D. (2024). Industrial Network Security: Securing critical infrastructure networks for smart grid, SCADA, and other Industrial Control Systems. Elsevier.
42. Kilovaty, I. (2019). Legally cognizable manipulation. Berkeley Tech. LJ, 34, 449.
43. Lepri, B., Oliver, N., & Pentland, A. (2021). Ethical machines: The human-centric use of artificial intelligence. ncbi.nlm.nih.gov
44. Lewis, T. G. (2019). Critical infrastructure protection in homeland security: defending a networked nation. John Wiley & Sons.
45. Manoharan, A., & Sarker, M. (2023). Revolutionizing Cybersecurity: Unleashing the Power of Artificial Intelligence and Machine Learning for Next-Generation Threat Detection. DOI: *https://www.doi.org/10.56726/IRJMETS32644*, 1.
46. Mallik, A. K. (2023). The future of the technology-based manufacturing in the European Union. Results in Engineering, 19, 101356.
47. Marin-Castro, H. M., & Tello-Leal, E. (2021). Event log preprocessing for process mining: a review. Applied Sciences, 11(22), 10556.

48. Metta, S., Chang, I., Parker, J., P. Roman, M., & F. Ehuan, A. (2024). Generative AI in Cybersecurity.
49. Mirsky, Y., Demontis, A., Kotak, J., Shankar, R., Gelei, D., Yang, L., Zhang, X., Lee, W., Elovici, Y., & Biggio, B. (2021). The Threat of Offensive AI to Organizations.
50. Molina, S. B., Nespoli, P., & Mármol, F. G. (2023). Tackling Cyberattacks through AI-based Reactive Systems: A Holistic Review and Future Vision. *arXiv preprint arXiv:2312.06229*.
51. Muth, J. S. (2023). The dazzled rope of lightning against the cloud is not the downward bolt but the compelled upstroke through the heated ether: stories.
52. Musser, M., Lohn, A., X. Dempsey, J., Spring, J., Shankar Siva Kumar, R., Leong, B., Liaghati, C., Martinez, C., D. Grant, C., Rohrer, D., Frase, H., Elliott, J., Bansemer, J., Rodriguez, M., Regan, M., Chowdhury, R., & Hermanek, S. (2023). Adversarial
53. Machine Learning and Cybersecurity: Risks, Challenges, and Legal Implications.
54. Nowrozy, R., & Jam, F. (2024). AI-Driven Security: Challenges and Opportunities. Journal of Cybersecurity and Information Assurance, 2(1), 1–12.
55. Nowrozy, R., & Jam, D. (2024). Embracing the Generative AI Revolution: Advancing Tertiary Education in Cybersecurity with GPT. arXiv preprint arXiv:2403.11402.
56. Opoku-Mensah, E., Abilimi, A. C., & Amoako, L. (2013). The Imperative Information Security Management System Measures In the Public Sectors of Ghana. A Case Study of the Ghana Audit Service. *International Journal on Computer Science and Engineering (IJCSE)*, 760-769.
57. Opoku-Mensah, E., Abilimi, C. A., & Boateng, F. O. (2013). Comparative analysis of efficiency of fibonacci random number generator algorithm and gaussian Random Number Generator Algorithm in a cryptographic system. *Comput. Eng. Intell. Syst*, *4*, 50-57.
58. Oseni, A., Moustafa, N., Janicke, H., Liu, P., Tari, Z., & Vasilakos, A. (2021). Security and Privacy for Artificial Intelligence: Opportunities and Challenges.
59. Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., ... & Xu, W. (2023). Six human-centered artificial intelligence grand challenges. International Journal of Human–Computer Interaction, 39(3), 391-437.
60. Parn, E. A., & Edwards, D. (2019). Cyber threats confronting the digital built environment: Common data environment vulnerabilities and block chain deterrence. Engineering, Construction and Architectural Management, 26(2), 245-266.
61. Pasha, S. A., Ali, S., & Jeljeli, R. (2022). Artificial intelligence implementation to counteract cybercrimes against children in Pakistan. *Human Arenas*, 1-19.
62. Radanliev, P., De Roure, D., Maple, C., & Ani, U. (2022). Super-forecasting the 'technological singularity' risks from artificial intelligence. *ncbi.nlm.nih.gov*
63. Raimundo, R. & Rosário, A. (2021). The Impact of Artificial Intelligence on Data System Security: A Literature Review. *ncbi.nlm.nih.gov*
64. Riggs, H., Tufail, S., Parvez, I., Tariq, M., Khan, M. A., Amir, A., ... & Sarwat, A. I. (2023). Impact, vulnerabilities, and mitigation strategies for cyber-secure critical infrastructure. Sensors, 23(8), 4060.
65. Rjoub, G., Bentahar, J., Abdel Wahab, O., Mizouni, R., Song, A., Cohen, R., Otrok, H., & Mourad, A. (2023). A Survey on Explainable Artificial Intelligence for Cybersecurity.
66. Sarker, I. H. (2022). AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. SN Computer Science, 3(2), 158.
67. Sarker, I. H. (2024). AI-Driven Cybersecurity and Threat Intelligence: Cyber Automation, Intelligent Decision-Making and Explainability. Springer Nature.
68. Sun, J., Li, Y., & Zhang, J. (2022). Adversarial Attacks on Machine Learning Models in Cybersecurity. IEEE Transactions on Information Forensics and Security, 17(1), 1–12.
69. Sun, L., Dou, Y., Yang, C., Zhang, K., Wang, J., Philip, S. Y., ... & Li, B. (2022). Adversarial attack and defense on graph data: A survey. IEEE Transactions on Knowledge and Data Engineering.

70. Wang, Y., Su, Z., Guo, S., Dai, M., Luan, T. H., & Liu, Y. (2023). A survey on digital twins: architecture, enabling technologies, security and privacy, and future prospects. IEEE Internet of Things Journal.
71. Xia, X., Bao, L., Lo, D., Xing, Z., Hassan, A. E., & Li, S. (2017). Measuring program comprehension: A large-scale field study with professionals. IEEE Transactions on Software Engineering, 44(10), 951-976.
72. Xu, H., Wu, J., Pan, Q., Guan, X., & Guizani, M. (2023). A survey on digital twin for industrial internet of things: Applications, technologies and tools. IEEE Communications Surveys & Tutorials.
73. Xu, T., & Zhou, Y. (2015). Systems approaches to tackling configuration errors: A survey. ACM Computing Surveys (CSUR), 47(4), 1-41.
74. Yeboah, T., Opoku-Mensah, E., & AyaabaAbilimi, C. (2013). A Proposed Multiple Scan Biometric-Based Registration System for Ghana Electoral Commission. *Journal of Engineering, Computers & Applied Sciences (JEC&AS)*, *2*(7).
75. Zaid, A., & Garai, A. (2024). Cybersecurity Threats in the Age of Artificial Intelligence. Journal of Cybersecurity and Information Assurance, 2(2), 1–10.
76. Zaid, T., & Garai, S. (2024). Emerging Trends in Cybersecurity: A Holistic View on Current Threats, Assessing Solutions, and Pioneering New Frontiers. Blockchain in Healthcare Today, 7.
77. Zarpelão, B. B., Miani, R. S., Kawakani, C. T., & De Alvarenga, S. C. (2017). A survey of intrusion detection in Internet of Things. Journal of Network and Computer Applications, 84, 25-37.
78. Zheng, Y., Li, Z., Xu, X., & Zhao, Q. (2022). Dynamic defenses in cyber security: Techniques, methods and challenges. Digital Communications and Networks, 8(4), 422-435.
79. Zeng, Y., Lin, H., Zhang, J., Yang, D., Jia, R., & Shi, W. (2024). How johnny can persuade llms to jailbreak them: Rethinking persuasion to challenge AI safety by humanizing llms. arXiv preprint arXiv:2401.06373.