American Journal of Science and Learning for Development

*Volume* **3**, *Issue* **7** | *July-2024*

**ISSN: 2835-2157**

# Risk Assessment of Artificial Intelligence Systems in Cybersecurity

**Arun Pandiyan Perumal[1], Pradeep Chintale[2], Ramasankar Molleti[3],Gopi Desaboyina[4]**

[1]Fremont, California, 94538,
[2]Lead Cloud Solution Engineer, SEI Investment Company, Downingtown, USA
[3]Address: 11227 Lost Maples Trl, Frisco, TX, 75035,
[4]Sr Systems Engineer, SEI Investment Company, Phoenixville , Pennsylvania, USA

**Abstract.**The article focuses on the vital function of AI (Artificial Intelligence) in cybersecurity measures and argues for effective risk assessment techniques in AI-powered cybersecurity. In this section, the article explored AI in cybersecurity by emphasizing the AI technologies that are in use and what their applications are in the field of threat detection, vulnerability management, and proactive defense mechanisms. Moreover, the article looked into the AI security risk types, such as malicious AI, biases, fairness issues, potential vulnerabilities, and the ethical questions the AI process causes. The paper explained the concepts of frameworks and methodologies for assessing risks in AI systems, beginning with existing risk assessment frameworks, such as the NIST cybersecurity framework and the FAIR framework methodologies. Risk mitigation strategies of AI systems, regulatory and ethical issues, and future AI and cybersecurity problems concerning technological progress are also assessed. As concluded, implementing regulations for compliance, ethical principles, and technological developments is the key to meeting the new challenges and developing safe and sustainable digital systems. The recommendations comprise fostering transparency, accountability, and continuous education and awareness programs to enable people to manage ethical dilemmas and mitigate the risks very well.

**Keywords:** cybersecurity, NIST, Risk mitigation, vulnerabilities, methodologies.

Citation: Hairul Huda, Siti Nursyamsiyah. Al Islam and Kemuhammadiyahan as Driving Force for Lecturer Performance at Universitas Muhammadiyah Jember. American Journal of Science and Learning for Development 2024, 3 (7), 49-60

Received: 14th June 2024
Revised: 15th July 2024
Accepted: 23th July 2024
Published: 29th July 2024

## 1. Introduction

Artificial intelligence (AI) is the main driving force of cybersecurity, bringing forth a new way of defending against future threats on the digital front [1]. AI that can automate tasks, detect threats, and enforce security is considered one of the principal technologies that defend against different types of cyberattacks [1]. The power of AI and cybersecurity is evidenced by the ever-growing number of digital threats that call for more advanced solutions for protection [2]. It requires a clear perception of AI roles in the cybersecurity field, its applications, limitations, and the trends that will arise in the future [2]. On the other hand, the importance of establishing rigorous risk assessment methodologies cannot be overstressed as AIs become incorporated into cybersecurity infrastructure because, being the same, they might create vulnerabilities and threats [3].

Being that AI-powered cybersecurity has a built-in form of risk assessment, there is a gap that needs to be closed across various application fields. Hence, there is a need to implement a structured approach to identification, valuation, and risk mitigation [3]. Consequently, this article aims to unfold the significance of AI in cybersecurity, illustrate the usage of risk assessment, and give the readers insights into methodologies for assessing risks linked to AI systems in the cybersecurity context.

## i. UNDERSTANDING ARTIFICIAL INTELLIGENCE IN CYBERSECURITY.

### *i. Overview of AI technologies used in cybersecurity*

The advancement of AI technology has been vital in enhancing cyber security through complex algorithms applied to combat emerging cybersecurity threats. It refers to a group of different technologies, such as machine learning, which are good at analyzing large quantities of data for identifying security breach patterns [4]. The real-time response to advanced cyberattacks is made more accessible by machine learning techniques such as support vector machines and random forest algorithms [5]. Moreover, AI-enabled cybersecurity systems could benefit from continual learning as the systems progress over time. As a result, the organization's overall safety is getting stronger [6]. Similarly, the effectiveness of the vulnerability management techniques is increased by locating possible countermeasures even within the present security measures and implementing proactive mitigation strategies that also consider the security risks [5]. Consequently, AI in cybersecurity is improving the process of dealing with cyber threats when they happen in real-time.

TABLE 1. TOP 10 USES OF AI FOR CYBERSECURITY [10].

| *Use of AI for Cybersecurity* | *Description* |
|---|---|
| **Identifies Unknown Threats** | Utilizes AI to effectively identify and prevent unknown threats, which could cause severe damage if undetected. |

| | |
|---|---|
| **Handles a Lot of Data** | Automatically scans and identifies disguised threats within vast network traffic, streamlining the detection process. |
| **Learns More Over Time** | Utilizes machine learning and deep learning techniques to analyze network behavior and enhance future security measures. |
| **Better Vulnerability Management** | Analyzes existing security measures to identify weak points, enabling businesses to focus on critical security tasks. |
| **Better Overall Security** | Detects and prioritizes all types of attacks, even when dealing with multiple threats simultaneously, reducing the risk of human error. |
| **Duplicative Processes Reduced** | Handles monotonous and repetitive security tasks regularly, ensuring network security best practices are implemented consistently. |
| **Securing Authentication** | Provides additional security layers using tools like facial recognition and CAPTCHA to prevent credential stuffing and brute force attacks. |

| | |
|---|---|
| **Eliminates Time-Consuming Tasks** | Scans vast data to identify potential threats and reduces false positives, allowing human experts to focus on more critical security tasks. |
| **Battling Bots** | Recognizes and blocks bots by identifying their patterns, creating secure captchas, and deploying honeypots to trap them, thereby enhancing network security. |

AI supports cybersecurity through its ability to detect unknown threats, process large volumes of data for clearer detection, and continuously learn from a system's behavior to improve security defenses. Furthermore, it enhances vulnerability management by examining existing security measures vulnerabilities, prioritizing security tasks, and faster detection and response of threats. AI reinforces authentication security, eliminates time-consuming tasks, and executes bots with the help of pattern recognition and countermeasures. These applications evidence the huge AI ability to reinforce network security in general and to deal adequately with various cyber threats, as demonstrated in Table 1 above.

i. *Examples of AI applications in cybersecurity.*

AI applications in cyber security cover several fields, such as intrusion detection, malware detection, and spam filtering. ML and DL techniques play an important role in current cyber defense schemes, adding sophisticated capabilities to detect and stop new cyber threat types [8]. ML techniques make it possible to filter out any abnormal pattern that could be a clue to data breaches. Equally, the DL algorithms enhance the detection of complex malware variants and deep cyberattacks [9]. Nevertheless, the rapid spread of AI in the cybersecurity field provokes issues with the opacity and lack of interpretability of AI-based defensive mechanisms [7]. The XAI methodologies are challenges resolved by enhancing the transparency of AI models, which allows security experts and end-users to understand and trust the decision-making algorithms used in cyber defense mechanisms [7]. As a result, the implementation of AI applications along with explanatory AI techniques is likely to reinforce cyber security measures without compromising on understandability and interpretability in AI-driven cyber defense strategies.

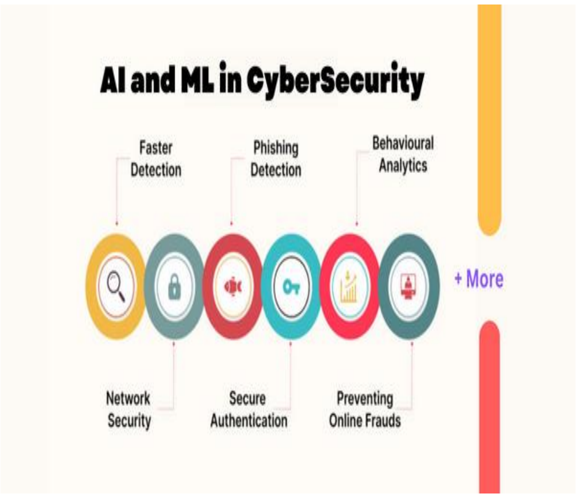

Fig 1: Ai and ML in Cybersecurity [16]

Fig. 1 visualizes the experience of adversarial machine learning and its impact on cybersecurity practices. Adversarial machine learning refers to training models to detect and classify adversarial factors, such as inputs carefully designed to mislead the models. In cybersecurity, intrusion detection is a crucial tool that assists in detecting and defending attacks meant to dodge security systems or misconstruing malware. With the knowledge of adversarial ML, cybersecurity experts can create indestructible defensive systems against many proficient cyberattacks.

### i. *Benefits of AI in Enhancing Cybersecurity Measures*

Artificial Intelligence (AI) is a critical factor that boosts cybersecurity capabilities with attention to threat detection, mitigation, and response. Machine learning algorithms on the basis of AI are capable of mass data analysis to detect patterns that might imply a cyber threat. It increases the efficiency and accuracy of intrusion detection systems [11]. Additionally, AI-based cybersecurity instruments exploit predictive techniques to predict and stop cyberattacks from causing a significant influence. It provides proactive defense mechanisms for developing threats such as ransomware, phishing, and malware [12]. Also, AI-empowered solutions such as automatic responses allow pressing matters to be addressed and ease the tasks of cybersecurity teams [13]. Overall, AI is a viable solution that reinforces defense mechanisms and allows organizations to be resilient while fighting against dynamic threat landscapes and protecting precious assets efficiently.

## i. TYPES OF RISKS ASSOCIATED WITH AI IN CYBERSECURITY.

### i. *Threats posed by malicious AI*

Today, cybercriminals are rapidly exploiting or misusing the capabilities provided by artificial intelligence (AI) to perform different kinds of cyber threats, such as integrity attacks, unexpected AI results, and algorithmic trading. Through targeting AI models' algorithms, threat actors aim to take advantage of vulnerabilities in cybersecurity [17].

Integrity attack is designed to mutilate the reliability of AI by tricking data inputs or algorithmic processes and thereby finding the system incapable of making the right decisions [14]. Likewise, unforeseen outcomes of AI can arise due to unanticipated interactions between various complex elements of artificial intelligence, thus giving rise to undesirable or negative consequences [15]. The use of automated weapons systems fed by AI is a new challenge that will require serious legislative measures as well as a great deal of cooperation between sectors for consideration.

### *i. Risks Associated with AI Bias and Fairness*

One of the risks that place AI technology applications at risk is that AI systems can suffer from biases already in the training data, algorithms, and decision-making processes, leading to discriminatory results [18]. It performs to support the social order and spoils the moral principles of honesty and equality. In reality, biased AI algorithms in hiring processes could be unfair to some minority groups, producing a system of discrimination [20]. Along with the ethical risks involved in these intelligent systems, the absence of transparency and accountability and, hence, the inability to deal with bias hinder their progress. Therefore, a comprehensive regulation and standards framework should be established to combat AI bias and the disparities in algorithm-based decision-making.

### *C. Potential Vulnerabilities in AI-Based Systems*

Artificial intelligence (AI) based systems introduce their unique vulnerabilities. They include data poisoning through adversarial attacks and lack of transparency, which threat actors can fraudulently exploit to compromise system integrity and security [22]. Data poisoning attacks are a construction of training data corrupted by AI models, which causes inaccurate or biased decisions [23]. Adversarial attacks take advantage of AI algorithms' susceptibility to malicious attacks by slight alterations in the input data and thus impair the systems' efficacy and degrade their trust [21]. AI is not built to be transparent about decision-making, which hinders practical vulnerability assessment and risk mitigation [21]. Therefore, the process necessitates a comprehensive approach that is based on thorough cyber safety procedures, frequent software testing, and continuous supervision of AI systems.

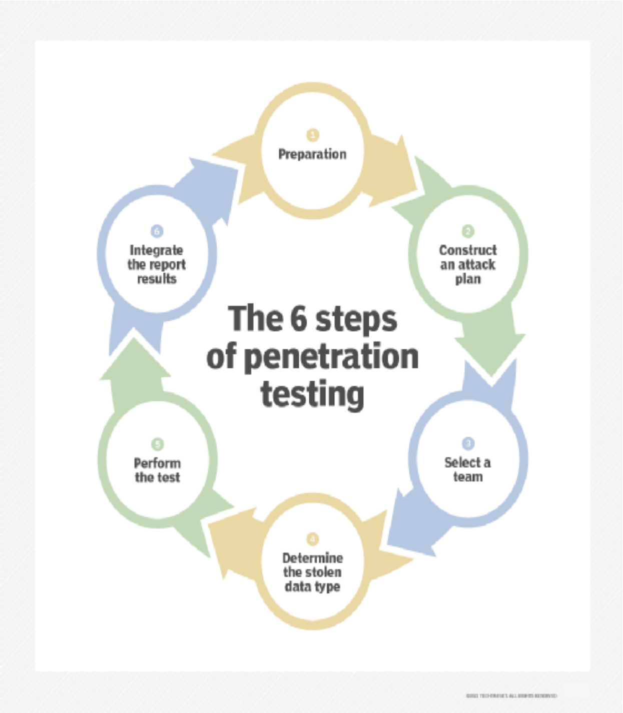
The 6 steps of penetration testing
1
Preparation
2
Construct an attack plan
3
Select a team
4
Determine the stolen data type
5
Perform the test
6
Integrate the report results
©2021 TECHTARGET. ALL RIGHTS RESERVED

Fig 2: Penetration Testing Process [TechTarget]

The fig. 2 depicts the procedure of penetration testing with several stages. Firstly, it starts with the reconnaissance, where the information on the target is collected. Next comes scanning where weaknesses of the system are identified. Next, the penetration tester attempts to gain access to the system using the exploitation technique. Once the tester is in the system, he/she will then elevate the privileges to determine the magnitude of potential damages. Finally, the report which details the findings and recommendations for improving the system's security posture is generated.

i. *Ethical problems about AI in cybersecurity arise.*

The fact that AI technologies are becoming very popular in cybersecurity poses inroads of ethical dilemmas such as algorithmic bias, data privacy, and accountability. AI-based cyber security systems should walk a tightrope when it comes to improving safety measures as well as protecting the rights and liberties of individuals [26]. On the other side, the use of AI to detect threats and the surveillance of citizens prompts questions about personal privacy and the possibility of misuse of personal data [24]. Moreover, the deceptiveness of artificial intelligence's decision-making processes has also resulted in the raising of questions on responsibility and transparency [19]. As such, it necessitates the formulation of ethical frameworks and regulations to ensure the responsible integration of AI in cybersecurity [27]. Hence, settling ethical dilemmas necessitates interdisciplinary collaboration and participation of stakeholders as a way of ensuring that AI technologies in cybersecurity hold ethical principles that mutually benefit both people and society.

## i. FRAMEWORKS FOR ASSESSING RISKS IN AI SYSTEMS.

### *i. Overview of existing risk assessment frameworks*

Evaluating risks like data privacy, robustness, fairness, and reliability is the definition of frameworks for AI systems risk assessment [28]. The dilemmas and conflicts are represented through these procedures such as the AI safety and efficiency [29]. Moreover, these approaches should be formed by multi-disciplinary groups involving legal, ethical and other social aspects of AI governance [30]. The judgement of ethics impact and social influence made by the DRESS-eAI where data point [29] is considered. These standards will thereby act as a guide to building public trust and ensure that the enterprises AI systems meet the ethical standard.

### i. *NIST Framework for Improving Critical Infrastructure Cybersecurity*

Security experts apply the MITRE Cybersecurity Criteria and NIST Cybersecurity Framework to assess the security risks of the AI systems [32]. Establishing business parameters to steer efforts and integrating cybersecurity risks into risk management is a part of the NIST Cybersecurity Framework for addressing cybersecurity issues [32]The plan is Identify, Protect, Detect, Respond, and Recovery process which consists of: identifying, protecting, detecting, responding, and recovering [33]. For instance, TTPs, which refers to strategies, techniques and procedures for addressing cyber threats and recovering from potentially malicious activities, are among the MITRE Cybersecurity Criteria [32]. Frameworks provide incredible opportunities to improve an organization's cybersecurity, minimize cyber threats and increase the resilience of the organization against constantly changing attacks. These systems positively contribute to the effectiveness of coordinated approaches to defending critical infrastructures and cyber-hazards.

### i. *FAIR (Factor Analysis of Information Risk) framework*

FAIR (Framework Factor Analysis of Information Risk) is used to perform cybersecurity risk assessment that is quantitative in nature. Organizations should be able to give security efforts a higher priority using FAIR model that shows the dollars spent on cyber-security over time [34]. On the other hand, some limitations such as limitations on statistical distribution and inability to improve the model structure by increasing complexity affect it [35]. To solve these, FAIR-BN is developed using Bayesian Networks along FAIR, providing capabilities to alter and expand the architecture [35]. Organizations are increasingly becoming subject to cyber security threats, and thus, adopting the FAIR framework is necessary to enhance their security postures [36]. The FAIR approach involves a structured methodology integrated into the strategic aim of fortifying an organization against cyber threats.

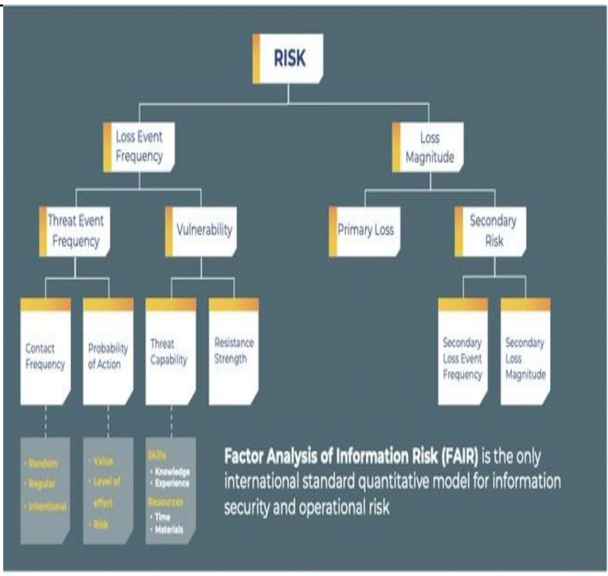

Fig 3: FAIR Analysis framework [39]

Fig. 3 stands for the risk taxonomy based on the FAIR methodology. It categorizes risks into six main areas: Frequency of Loss Event (LEF), Frequency of Threat Event (TEF), Vulnerability, Loss Magnitude (LM), Threat Capability (TCap), and Control Strength (CS). The former - LEF – stands for the rate of loss events, and the latter - TEF – means threat events. Assessment of vulnerability understands the system weaknesses, while LM stands for loss magnitude evaluation. TCap indicates the enemy's skills, while CS implies the ability of existing controls to mitigate the risks.

i. *MITRE ATT&CK framework*

The MITRE ATT&CK framework provides a thorough knowledge of adversary methods, techniques, and procedures. It has several functions, such as blue teaming, red teaming, and security operation center (SOC) maturity assessment. Nevertheless, existing research commonly emphasizes particular attack forms or organizational culture, leaving a hole linking both aspects mutually [37]. The breach has been addressed by relating organizational & individual culture factors with security vulnerabilities, as mapped by the MITRE ATT&CK framework [38]. Moreover, advanced threat modeling languages, like those based on the MITRE Enterprise ATT&CK Matrix, will contribute to proactively solving the security problems in complex enterprise systems over the long run [40]. These languages involve simulating attacks and being able to perform the analysis of security settings and architectural changes for improved system security [40]. Consequently, the ATT&CK framework of MITRE is constantly developing to provide better insight into cyber threats and improve security readiness across varied domains.

i. METHODOLOGIES FOR CONDUCTING RISK ASSESSMENTS IN AI SYSTEMS.

The risk assessment method of AI systems consists of several stages, such as identifying assets and threats. The time is for constructing all AI system components including data, algorithms, and system infrastructure, followed by assessing all risks and vulnerable factors that may influence the security and performance of AI systems [41] After the evaluation, the assets will be linked to their related vulnerabilities. As such, it outlines the vulnerabilities and exploits and the data misuse or outdated software components. Subsequently, the outcome analysis is carried out to determine the probable effects of successful cyber-attacks on the AI system [44]. Failure factors of analysis involve data loss, system inactivity, and reputation loss, which provide good details about the risks detected. Therefore, probability evaluation of the risk occurrence will be done by analyzing the probability of each identified risk [43]. The approach entails evaluating historical data, intelligence information, and the consultation of subject matter experts to determine the level of various threats occurring. More precise and personalized approaches like fuzzy and neural network models are developed for risk assessment [42]. These risk analysis tools are not limited to the conventional sectors. They are also used in other sectors, such as cyber-physical power systems and liquefied natural gas. Such courses demonstrate risk assessment techniques' crucial role in avoiding critical infrastructure and emerging technology risks.

### i. BEST PRACTICES FOR MITIGATING RISKS IN AI SYSTEMS

AI risk mitigation best practices require developing cutting-edge security systems that protect against various risks. The target is to use encryption protocols, access controls, and secure development approaches to safeguard AI systems from unauthorized access and data breaches [45]. However, the other necessary process is auditing AI systems so they can be exposed to where the attackers can abuse them [46]. Companies can use full-scale and professional tests like penetration tests and vulnerability scanning to scan for any existing weaknesses and fix them before hackers can take advantage of them. Algorithm transparency and accessibility to AI systems are the following essential things to be addressed [47]. AI companies build trust and responsibility towards their users and stakeholders through the transparent descriptions of AI-driven algorithms. This can be achieved by cutting the probability of undesired consequences. Staff training and awareness campaigns about cybersecurity are also indispensable to ensure that teams have the required knowledge and skills for identifying and managing AI risks [48]. Businesses may prevent cyber threats and secure their AI systems by allocating resources to training and human capital development. Ultimately, the ideal practices are the basis for firm and secure AI systems that successfully list and remove several perils in an ever-expanding and complicated digital arena.

### i. REGULATORY AND ETHICAL CONSIDERATIONS.

Presently, regulatory and ethical considerations have been vital in developing and implementing AI systems, especially in the cybersecurity field. Following data protection laws, such as the GDPR and the CCPA, is necessary to ensure the data security of individuals and their rights [49]. Data collection, processing, and storage rules inform organizations what to do so that the dangers to privacy do not overtake the benefits of AI [50]. However, more ethical rules must be developed for discrimination, transparency, and accountability [51]. Organizations have to establish an ethical process for creating and using AI systems that will be ethical, transparent, and with people's right to privacy considered. In addition, being responsible and accountable when managing AI-driven cyber security is also crucial to keep malicious attacks away and maintain the credibility and trustworthiness of AI-based security systems [52]. Through the combination of compliance requirements and ethical principles in AI systems development and use, organizations can be at the forefront of AI ethical adoption and the increase in the avoidance of risks and ethical problems.

## i. FUTURE TRENDS AND CHALLENGES.

Forecasts of AI and cybersecurity trends include new technologies and challenges that will transform the cybersecurity landscape. AI and machine learning will change autonomic computing by improving resource autonomy and performance. However, there are still a few challenges in attaining complete automation and scalability [53]. Moreover, e-waste management (WEEE) is a double-edged sword, environmentally and legally, so innovative methods of waste disposal and recycling must be found [54]. Smart cities depend heavily on technological developments such as IoT, big data, and AI to keep pace with urbanization trends and enhance the quality of life for citizens [54]. Some expected challenges are safeguarding data privacy and security in the smart city infrastructure, as well as addressing technological restrictions and regulatory frameworks. Strategies to address the emerging trends and challenges include employing circular principles, enhancing regulatory enforcement, and promoting technological innovations in cybersecurity as well as waste management. Through the solution of these problems and the use of the new technology, participants can create sustainable and secure digital ecosystems for the future.

## CONCLUSION.

In conclusion, AI integration provides the best service for both opportunities and challenges in cybersecurity. The continuous risk assessment with AI-driven cybersecurity is thus of the essence, as AI technologies are actively developed and the threats in cyber security become more sophisticated. Companies should consider risk management as a critical aspect of AI systems security by building protection against probable problems and threats. Some of the main issues touched upon during this discussion revolve around the immense contribution of AI towards the formation of robust cybersecurity defenses, the use of AI in different areas of cybersecurity, such as malware detection and threat prevention, and risks posed by AI systems, such as malicious AI, biases, and security flaws. Furthermore, the norms and techniques for analyzing risks of AI systems can serve as a practical guide for companies to determine, assess, and eliminate risks successfully. The responsible development and deployment of AI systems in information security will be ensured through regulatory and ethics oversights, and the trends should focus on innovations and adaptations to counter evolving threats and technologies. Consequently, appropriate steps that emphasize risk management and adopting secure data storage and protection strategies must be taken for organizations to cope effectively with the dynamic landscape of AI-powered cybersecurity.

REFERENCES.

1. Kaur, R., Gabrijelčič, D., & Klobučar, T. (2023). "Artificial Intelligence for Cybersecurity: Literature Review and Future Research Directions." Information Fusion, vol. 101804.
2. Aslam, M. (2024). "AI and Cybersecurity: An Ever-Evolving Landscape." International Journal of Advanced Engineering Technologies and Innovations, vol. 1, no. 1, pp. 52-71.
3. Wiafe, F., Koranteng, N., Obeng, E. N., Assyne, N., Wiafe, A., & Gulliver, S. R. (2020). "Artificial Intelligence for Cybersecurity: A Systematic Mapping of Literature." IEEE Access, vol. 8, pp. 146598-146612.
4. Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A., Aziz, N., Capretz, L. F., & Abdulkadir, S. J. (2022). "Detecting Cybersecurity Attacks in Internet of Things Using Artificial Intelligence Methods: A Systematic Literature Review." Electronics, vol. 11, no. 2, p. 198.
5. Dash, B., Ansari, M. F., Sharma, P., & Ali, A. (2022). "Threats and Opportunities with AI-Based Cyber Security Intrusion Detection: A Review." International Journal of Software Engineering & Applications (IJSEA), vol. 13, no. 5.
6. Jain, V., Balakrishnan, A., Chintale, P., Gadiparthi, S., & Najana, M. (2024). "Blockchain Empowerment in Sanctions and AML Compliance: A Transparent Approach." International Journal of Computer Trends and Technology, vol. 72, no. 5, pp. 1-10.
7. Zeadally, S., Adi, E., Baig, Z., & Khan, I. A. (2020). "Harnessing Artificial Intelligence Capabilities to Improve Cybersecurity." IEEE Access, vol. 8, pp. 23817-23837.
8. Soni, V. D. (2020). "Challenges and Solution for Artificial Intelligence in Cybersecurity of the USA." SSRN, 3624487.
9. Naik, B., Mehta, A., Yagnik, H., & Shah, M. (2022). "The Impacts of Artificial Intelligence Techniques in Augmentation of Cybersecurity: A Comprehensive Review." Complex & Intelligent Systems, vol. 8, no. 2, pp. 1763-1780.
10. Nadella, G. S., & Gonaygunta, H. (2021). "Enhancing Cybersecurity with Artificial Intelligence: Predictive Techniques and Challenges in the Age of IoT."
11. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). "A Survey on Bias and Fairness in Machine Learning." ACM Computing Surveys (CSUR), vol. 54, no. 6, pp. 1-35.
12. Landers, R. N., & Behrend, T. S. (2023). "Auditing the AI Auditors: A Framework for Evaluating Fairness and Bias in High Stakes AI Predictive Models." American Psychologist, vol. 78, no. 1, p. 36.
13. Fletcher, R., Nakeshimana, R. A., & Olubeko, O. (2021). "Addressing Fairness, Bias, and Appropriate Use of Artificial Intelligence and Machine Learning in Global Health." Frontiers in Artificial Intelligence, vol. 3, p. 561802.
14. Berghoff, M., Neu, J., & von Twickel, A. (2020). "Vulnerabilities of Connectionist AI Applications: Evaluation and Defense." Frontiers in Big Data, vol. 3, p. 23.
15. Breda, P., Markova, R., Abdin, A., Jha, D., Carlo, A., & Mantı, N. P. (2022). "Cyber Vulnerabilities and Risks of AI Technologies in Space Applications." In 73rd International Astronautical

Congress (IAC), Paris, France.