Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 9
AIDML 2025

Applications and Challenges of Artificial Intelligence in the Field of Cybersecurity

Qiyue Yang *

School of Computer and Cyber Sciences, Communication University of China, Beijing, 100024, China

* Corresponding Author Email: ang7ue@gmail.com

Abstract. With the rapid development and widespread application of computer technologies, cyber-attacks have become increasingly diverse and intelligent. Traditional cyber defense methods are no longer sufficient to meet current security demands. Therefore, introducing more efficient and intelligent protection technologies has become critically important. This paper focuses on the applications and challenges of artificial intelligence (AI) in the field of cybersecurity. Through a combination of literature review and case analysis, it systematically examines the application of AI in areas such as intrusion detection, malware identification, vulnerability discovery and risk assessment, threat intelligence analysis, and data center security. Furthermore, it evaluates the challenges AI faces in terms of interpretability, data quality, implementation costs, interdisciplinary collaboration, privacy protection, and system vulnerabilities. The findings indicate that AI technologies are widely adopted in the cybersecurity domain, significantly advancing the development of security techniques. Looking ahead, further efforts in technology, systems, talent development, and policy coordination are required to build intelligent security frameworks and to promote the deeper integration of AI in cybersecurity.

Keywords: Artificial Intelligence (AI); Cybersecurity; Cyber-attack; Machine Learning (ML).

1. Introduction

With the rapid development of global information technology, cybersecurity has become an important research topic. Individuals, enterprises and even governments are now facing the threat of cyber-attacks. The constantly developing and complex means of attack pose a serious challenge to the whole society and country. Traditional defense methods have made it difficult to meet the needs of today's cybersecurity, and there is an urgent need to introduce more efficient and intelligent protection technology.

In recent years, the breakthrough of AI technology in the fields of big data analysis, machine learning (ML) and deep learning has brought new development opportunities for the field of cybersecurity. The use of AI technology can realize the real-time monitoring of massive network traffic, the accurate identification of threat means and the rapid response to abnormal behavior, thus greatly improving the defense efficiency of intrusion detection, vulnerability analysis and anti-fraud. At the same time, the adaptive learning mechanism of AI technology enables the security defense system to evolve continuously, actively respond to new attack means, and further enhance the overall protection level of cybersecurity.

However, with the wide application of AI technology in cybersecurity, related problems are gradually emerging. Data quality and privacy protection, adversarial attacks, false positives and false judgments, and attacks against large models have become important bottlenecks restricting AI security applications.

This paper will gradually analyze the specific application and advantages of AI in the field of cybersecurity, as well as the challenges and shortcomings faced by the current technology, and realize the systematic research of AI technology in cybersecurity, so as to provide a theoretical basis and reference for the construction of a more intelligent, efficient and secure cybersecurity system.

2. Applications of AI in Cybersecurity

2.1. Intrusion Detection and Defense

The application of AI in cybersecurity, especially in the field of intrusion detection and defense, is increasingly becoming the core technology for building a powerful protection system. By analyzing vast amounts of network traffic data, large models can learn and capture the characteristics and patterns of cyber-attacks, enabling the accurate identification and defense against various types of intrusion activities, such as unauthorized access, data theft, system damage, distributed denial-of-service (DDoS) attacks, phishing attacks, and malware attacks [1]. This detection method based on an intelligent algorithm can not only help cybersecurity professionals better understand the nature of the attack, but also significantly improve defense efficiency, forming a powerful supplement to the traditional security means [1].

At the same time, in daily network usage, many professional cybersecurity software programs have begun to incorporate AI technologies to achieve intelligent intrusion detection. By expanding the detection scope and conducting real-time analysis of high-quality network data, such software can process and analyze information more efficiently, enabling quicker responses to potential security threats [2]. The deep integration of AI and intrusion detection technology makes it possible to intelligently monitor viruses and other abnormal behaviors, thus building more comprehensive and reliable protection support for the system and related equipment [2].

Further, with the help of AI technology, the defense task against network intrusion can also be significantly simplified and strengthened. For example, the network firewall embedded with AI technology performs well in preventing unauthorized access. By collecting and analyzing network data in real time, it can quickly identify and prevent various attacks and ensure the security of system information [3]. It is this ability to continuously acquire large volumes of information, accurately detect threats, and intelligently block cyber-attacks that enables AI-powered intrusion detection and defense systems to serve as a robust safeguard for cybersecurity, offering maximum protection for individuals and organizations alike.

2.2. Malware Detection Techniques

Malware remains one of the most significant cybersecurity threats today, and effective defense requires rapid and efficient analysis of the increasing volume of malware [4]. Traditional malware detection technologies are mainly divided into two types: static detection and dynamic detection [4]. The static detection method identifies malware by analyzing its binary files or the code instructions obtained through decompilation, without executing the code. However, this approach is susceptible to evasion techniques, such as obfuscation and syntax merging. In contrast, dynamic detection involves executing malware in a controlled environment and monitoring its behavior. While both techniques have their own advantages, manually analyzing each malware sample is time-consuming and resource-intensive in practice, which limits their effectiveness in large-scale applications [4].

To solve this dilemma, many malware detection systems based on AI emerge as the times require, especially those using deep learning algorithms. Compared with traditional methods, they not only perform better in detection accuracy, but also use resources more efficiently, significantly improving detection speed and overall performance [4]. Given the opacity of neural networks, many researchers have introduced various explainable AI (XAI) techniques to enhance the interpretability and transparency of AI-based malware detection systems. This helps ensure that such detection tools can maintain stable and efficient performance when deployed in new environments. Through the training model, these AI systems can automatically extract the characteristics of malicious software, and quickly process and analyze massive data, so as to achieve accurate detection of malicious software such as viruses, Trojans and worms.

At the same time, phishing, as another common means of cyber-attack, also threatens the security of users' sensitive information. Traditional phishing attacks use forged e-mail or false websites to induce

users to click links or download executable files with the help of human weakness, thus triggering the spread of malware or the disclosure of sensitive data [5]. In order to avoid this attack, users usually need to evaluate the legitimacy of links, which often depends on professional judgment on web code and security features. To address this issue, researchers have proposed AI-based methods utilizing Support Vector Machines (SVM), which effectively detect fraudulent banking websites by analyzing features such as IP addresses, SSL certificates, the number of dots in the URL, URL length, and the presence of blacklist keywords [5]. These methods achieved an accuracy of 98.86%, demonstrating that AI training can significantly compensate for the limitations of human awareness in cybersecurity [5].

Attackers also use the characteristics of JavaScript widely used in modern websites and online social media to spread malware by inserting malicious code, implementing phishing or triggering a passing download attack [5]. Faced with this situation, traditional detection methods often require advanced coding knowledge, which makes it difficult for ordinary users to identify malicious websites. To this end, AI technology has been employed to perform an in-depth analysis of JavaScript code characteristics—such as word length, character distribution, bytecode frequency, comment styles, and sensitive function calls—in order to detect obfuscated malicious JavaScript. This approach also provides a fail-safe mechanism to prevent the further spread of malware following a phishing attack [5]. In this way, AI technology not only strengthens the detection ability of malware and phishing attacks, but also further improves the overall cyber defense level, forming a more perfect and efficient security protection system.

2.3. Vulnerability Discovery and Risk Assessment

The application of the big model in cyber security is not only reflected in intrusion and malware detection but also extended to vulnerability analysis. Through deep mining of a large number of vulnerability data, these models can capture the characteristics and patterns of vulnerabilities, so as to accurately identify and evaluate various vulnerabilities in network systems, including security vulnerabilities in operating systems, applications and network protocols. With this method, cyber security professionals can more comprehensively understand the nature of vulnerabilities, and provide theoretical basis and data support for developing more efficient vulnerability repair strategies [1]. In addition, considering that in the past, hackers often used the weakness of slow response in vulnerability management to launch attacks, using AI technology to manage vulnerability databases can realize real-time reporting of attack attempts, thus significantly improving the security of the whole system.

On this basis, the generative language model plays an important role in identifying vulnerabilities and risk assessment. For example, ChatGPT was used to test code security and function related issues in the study. The model can detect potential buffer overflow risks in a given code fragment and explain in detail how to trigger the vulnerability [6]. Furthermore, ChatGPT also proposed a simpler solution. Although this solution is only for the code provided by the researchers, it can provide further optimization strategies when getting more user prompts [6]. In addition, ChatGPT also successfully found a vulnerability in the TLS protocol code extension, which may allow attackers to disclose sensitive information from the server memory; At the same time, it gives a detailed and simplified explanation of the source code for verification of Bitcoin, and assesses the risk of blockchain attack [6].

2.4. Threat Intelligence and Predictive Analysis

In the field of cybersecurity, threat intelligence processing using large models has become a cutting-edge application direction. Through the in-depth mining of massive threat intelligence data, these models can extract the inherent characteristics and patterns of threats, so as to accurately identify and evaluate cyber-attacks, malware, security vulnerabilities and other threats.

On the other hand, threat intelligence involves the systematic collection, in-depth analysis and efficient dissemination of potential or existing cyber threat information, and plays an important role in providing organizations with key knowledge and data needed to respond to cyber-attacks [7]. In recent years, AI and ML technologies have been widely integrated into the threat intelligence system, realizing the automatic collection, analysis and transmission of massive data [7]. Among them, threat hunting, as an important application of AI and ML in threat intelligence, identifies intrusion signs and abnormal activities in network infrastructure through active detection methods, thus significantly improving the efficiency and accuracy of threat detection [7].

In addition, these technical tools cover text data analysis technologies, including natural language processing, deep learning algorithms for pattern recognition in complex data structures, and graphic analysis methods for revealing complex relationships among network entities [7]. The AI and ML driven threat search platform can automatically filter and integrate a large amount of data, capture and mark complex threat patterns with errors, and its detection performance and accuracy far exceed the traditional manual detection method [7]. At the same time, the threat prediction system uses AI and ML to conduct in-depth analysis of historical data and identify potential attack patterns in advance, so that organizations can actively deploy defense measures before attacks occur to enhance overall security protection [7]. Finally, in terms of threat intelligence sharing, relevant technologies support the automatic integration and dissemination of intelligence data from multiple channels, effectively promoting information exchange and collaborative protection across organizations [7].

2.5. Security Applications in Data Centers

In many areas with high cybersecurity requirements, the data center is undoubtedly one of the most critical components. A major advantage of AI is that it can automate the operation and management process of the data center, thus effectively reducing the mistakes caused by human operations. In the data center, key indicators such as power consumption, bandwidth utilization and temperature are always under strict monitoring. With the fine management of AI technology, the operating efficiency of these indicators can be significantly improved.

In addition, the economic cost of hardware maintenance must be fully considered in the process of centralized management of the data center using the AI system. Because the data center carries important information of customers or organizations, its security must not only resist various cyber-attacks, but also against risks brought by external environmental factors. Therefore, ensuring the stability and security of equipment has become a basic task in data center management. In recent years, more and more enterprises and institutions have chosen to incorporate the AI system into the management system of the data center, in order to further improve the operational efficiency while improving the security protection capability. This trend fully reflects the important role of AI in the field of cybersecurity.

3. Challenges of AI in Cybersecurity

3.1. Explainability and Trustworthiness

In the process of applying AI technology to cybersecurity, an important challenge that needs urgent attention is the transparency problem caused by its "black box" nature. The so-called "black box" refers to the opacity of highly complex AI models, especially deep neural networks and large language models, in the decision-making mechanism, which makes it difficult for external observers to understand the basis of their output [8]. This lack of interpretability may weaken the user's trust in the system and affect the acceptance and effectiveness of AI aided decision-making in scenarios such as cybersecurity, which require high controllability and censorship.

Furthermore, the "hallucination" problem shown by the current mainstream large language models when generating text, that is, although the output content of the model seems reasonable, it may actually be inconsistent with the facts, which constitutes another key risk [1]. The existence of such

errors may lead to misjudgment and even security accidents when facing sensitive tasks such as cybersecurity. Therefore, although AI technology can provide auxiliary support for cybersecurity, its generated results should be regarded as a reference rather than the only decision-making basis.

3.2. Data Quality and Implementation Costs

When developing and deploying large-scale AI models in the field of cybersecurity, one of the core challenges lies in the lack of high-quality training data. The effectiveness of such models heavily relies on access to rich, well-structured datasets—such as threat intelligence databases, attack case repositories, and data provided by security vendors. However, these data sources often suffer from issues such as varying standards, inconsistent formats, and high noise levels, which can lead to misjudgments and performance degradation during the data cleaning and modeling phases [1]. Moreover, huge training data often involves a large amount of sensitive information, which brings additional security risks in the process of data collection and storage, increasing the vulnerability of the system [9].

At the same time, the complex architecture of the large model itself also causes it to consume a lot of computing resources in the process of training and tuning. The scale of model parameters is huge, involving multiple dimensions such as layers, attention mechanisms, and hidden units [1]. Each parameter debugging is accompanied by significant time, energy consumption and hardware cost, which makes the model optimization process have a high threshold in engineering practice.

To sum up, in the whole process from data acquisition to model training and deployment, AI applications in the field of cybersecurity face multi-dimensional challenges such as scarce data resources, unstable quality and high computing costs.

3.3. Talent Gap and Interdisciplinary Barriers

At present, in the process of building a vertical big data model for information security, the collaboration between AI technicians and cybersecurity professionals faces many challenges. This divide stems from AI specialists' mastery of advanced algorithms but limited insight into security architectures, while cybersecurity experts possess deep threat expertise yet lack proficiency in large-scale data modeling and algorithmic optimization [1].

In the big data environment, the cybersecurity task itself is highly professional and real-time, which puts forward higher requirements for AI technology. Researchers not only to master the abstract logic of the algorithm itself, but also to understand the background knowledge and potential logic relationship of security events. Therefore, a single technical team is often not competent for the design and optimization of the overall system.

In order to solve the above problems, it is urgent to build a collaborative mechanism based on interdisciplinary integration, which organically combines the professional capabilities of AI experts and cybersecurity experts [1]. Further, promotes the formation of a systematic cross-border cooperation framework and training system, enables the AI team to master basic security knowledge, and promotes the cybersecurity team to understand the principles and limitations of AI modeling, so as to propose more practical and robust modeling methods.

3.4. Privacy Concerns and Data Security

AI systems depend heavily on large-scale, multi-source datasets—such as user-generated content on social media—that often include substantial amounts of personally identifiable information. Moreover, every interaction with these systems produces message content, device metadata, log entries, and cookie data, all of which are automatically harvested and stored. Establishing robust governance over how this data is collected, shared, and retained is therefore crucial to prevent mishandling that could violate individual privacy or expose sensitive information to unauthorized parties or malicious actors.

At the same time, many AI tools also rely on third-party services to track users' online behavior, which increases the risk of data leakage on the one hand, and often lacks transparency when sharing data across platforms or institutions on the other hand. To this end, it is necessary to clearly define the identity of the third-party service provider and the specific responsibilities that should be undertaken in the process of data sharing and use, so as to ensure that all relevant stakeholders can fully understand the system's data management practices. Specifically, the system should make a detailed description of the collection, utilization, sharing and destruction of data throughout the life cycle, and clarify the division of responsibilities in the privacy policy to improve the openness and transparency of data operations.

3.5. System Vulnerabilities and Risk of Misuse

The essence of AI system is a set of computer programs developed by human beings, whose operation depends on preset logic instructions and protocols, which means that it can be modified and manipulated by individuals or groups with corresponding technical capabilities. In recent years, new cyber-attack tools based on large language models have emerged endlessly. Among them, PentestGPT has achieved full automation of the penetration test process with the help of ChatGPT technology, and can complete information collection, vulnerability scanning and utilization ring without intervention [10]. WormGPT can automatically generate code for launching various cyber-attacks through training and debugging a large number of malicious scripts, and has the ability to quickly upgrade and iterate [10]. In addition, FraudGPT synthesizes high simulation phishing emails based on a deep learning model. Its content and format can easily bypass the security detection of traditional email gateways, thus greatly improving the success rate of attacks [10]. Once the system is maliciously tampered with, the tools originally used to defend against network threats may be transformed into a means of attack itself. This potential controllability is one of the most concerning risks of AI in cybersecurity applications.

In addition, with the continuous evolution of malware and attack means, if the AI system does not update the model in time, it will also face the risk of identification failure. What is more serious is that the development speed of cybercrime often exceeds the update speed of defense technology. Therefore, the application of AI in cybersecurity must be supplemented by flexible and diverse functional design, and combined with the experience of professional security personnel to enhance the system's response capability in complex scenarios [5].

4. Conclusion

With the rapid development of AI technology, its application potential and practical effect in the field of cybersecurity are increasingly prominent. This study systematically combs the key application scenarios and challenges of AI in cybersecurity, and draws the following main findings and conclusions. First, this paper comprehensively discusses its application value in intrusion detection and defense, malware identification, vulnerability mining and risk assessment, threat intelligence analysis, and data center security protection. Especially in dealing with complex cyber-attacks, dynamic threat evolution and large-scale log data, AI technology provides strong support for building a more intelligent and active security protection system by virtue of its efficient learning ability and rapid response ability. For example, the performance of deep learning in abnormal traffic identification and malware detection has significantly exceeded the traditional rule-based methods, reflecting the breakthrough progress of AI in the dimension of security protection. The generative model also shows the potential to assist decision-making in automated penetration testing and code audit, and can quickly generate test scripts and propose repair suggestions. The AI driven threat hunting and prediction platform realizes active identification and early warning of emerging attack means through natural language processing, graph analysis, timing prediction and other technologies.

However, with the deepening of application, many challenges faced by AI in the field of cybersecurity have emerged gradually. This paper focuses on five levels of challenges: first, the interpretability and

credibility of the model. The current "black box" AI model lacks a transparent decision-making mechanism and cannot guarantee credibility; Secondly, data quality and implementation costs. Data in the field of cybersecurity is characterized by complexity, heterogeneity and high noise. However, high-quality annotation data is scarce, and a large model training process is accompanied by high computing and labor costs; Thirdly, there are knowledge barriers between AI and cybersecurity, and it is difficult for experts in different fields to collaborate, which hinders the realization of interdisciplinary integration and efficient collaboration; The fourth is privacy protection and data security. In the process of collecting and processing large amounts of data, AI needs to consider how to reasonably collect, share and store these data to prevent privacy disclosure; Finally, the big model itself also has the risk of being attacked and abused. For example, WormGPT, FraudGPT and other cases that have been hacked for automated attacks have posed new challenges to the existing security defense system.

To sum up, the trend of AI technology enabling cybersecurity has been irreversible, significantly improving the defense and response capabilities of cybersecurity. However, in order to truly realize its wide and reliable application, still need to continue to make efforts in the aspects of model interpretability, cross professional talent training, safety protection mechanism design, etc. Future research should build a more solid technical defense line for cyberspace security by building a controllable, credible and auditable intelligent security system, taking AI system as an auxiliary means, and relying on transparent regulatory mechanisms and multi-party collaboration.

References

[1] Hou Chao, Miao Haoyu, Bao Tianyuan, et al. Qiantan Da Moxing Zai Wangluo Anquan Zhong De Yingyong. Network Security and Informatization, 2025, (03): 1 - 3.

[2] Wang Tianyu, Guo Ying. Wangluo Anquan Fangyu Zhong Rengong Zhinen Jishu Yingyong Fenxi. Northeast Electric Power Technology, 2025, 46 (02): 40 - 42.

[3] Ansari M F, Dash B, Sharma P, et al. The Impact and Limitations of AI in Cybersecurity: A Literature Review. International Journal of Advanced Research in Computer and Communication Engineering, 2022.

[4] Zhang Z, Al Hamadi H, Damiani E, et al. Explainable AI Applications in Cyber Security: State-of-the-Art in Research. IEEE Access, 2022, 10: 93104 - 93139.

[5] Zeadally S, Adi E, Baig Z, et al. Harnessing AI Capabilities to Improve Cybersecurity. IEEE Access, 2020, 8: 23817 - 23837.

[6] Al-Hawawreh M, Aljuhani A, Jararweh Y. ChatGPT for Cybersecurity: Practical Applications, Challenges, and Future Directions. Cluster Computing, 2023, 26 (6): 3421 - 3436.

[7] Mohamed N. Current Trends in AI and ML for Cybersecurity: A State-of-the-Art Survey. Cogent Engineering, 2023, 10(2).

[8] Sontan A D, Samuel S V. The Intersection of AI and Cybersecurity: Challenges and Opportunities. World Journal of Advanced Research and Reviews, 2024, 21 (2): 1720 - 1736.

[9] Roshanaei M, Khan M R, Sylvester N N. Enhancing Cybersecurity Through AI and ML: Strategies, Challenges, and Future Directions. Journal of Information Security, 2024, 15 (3): 320 - 339.

[10] Yuan Weiguo, Zhang Xinyue, Yuchi Xuebiao. Shengchengshi Rengong Zhinen Jishu Dui Wangluo Anquan Lingyu De Yingxiang Fenxi Yu Qishi Jianyi. Information and Communication Technology and Policy, 2025, 51 (01): 2 - 9.