Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/47/20241480

Application and challenges of artificial intelligence in cybersecurity

Fangshu Li

School of Electronic Engineering and Computer Science, Queen Mary University of London, London, United Kingdom

ec21780@qmul.ac.uk

Abstract. This paper delves into the multi-faceted role of Artificial Intelligence (AI) in the field of cybersecurity. It conducts an in-depth analysis of two common security problems, spam email and DDos/Dos attacks. It focuses on the application of AI in resolving these problems and further on Event Management (SIEM) and Intrusion Detection Systems (IDS). Furthermore, this article outlines and discusses five key challenges that AI may encounter throughout its development, including issues related to reliability, unidentified threats, data privacy, ethical considerations, and the importance of an explainable AI. The paper underscores the remarkable potential of AI to improve our defences against emerging cybersecurity threats. It highlights the growing importance of addressing ethical and legal issues in the evolving technological environment. Finally, the article offers a comprehensive summary and valuable insights into the future prospects of AI in the realm of cybersecurity, confirming its crucial role in shaping the security landscape in the digital era.

Keywords: Artificial intelligence, Cybersecurity, Cyber Attack, Computer Security.

1. Introduction

Computer security is the process of protecting computer systems from potential threats. With the growing number of people using the Internet and electronic devices, there is a growing concern for the protection of private information and device systems from potential threats and attacks. Such hazards may arise from the internet or injected from a USB Driver. Cyber attacks are more frequent and widespread than attacks that involve direct physical insertion into a computer system. Therefore, discussions of computer security usually center around network security and gradually expand into more specific areas.

In the early stages of cyber attacks and hacking, the techniques were relatively basic and uncomplicated. However, as technology has advanced, hacking has become progressively frequent, more complex and harder to resolve. Managing large amounts of data manually is a time-consuming and inefficient process, which can also be laborious and draining for the workers. Consequently, automated detection and defence techniques, referred to as Artificial Intelligence (AI), have become crucial [1].

The two commonly cited methods in AI are deep learning and machine learning. The accuracy of the results depends on the number of trained datasets used and is generally reliable, though there is still scope for enhancement. It is impossible to guarantee that the inaccuracies will not cause significant

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/47/20241480

cybersecurity issues. Hence, AI is expected to be transparent, reliable, and trustworthy. AI techniques have been extensively used in the cybersecurity sector. This article analyzes the common applications of AI in security and its limitations in the past 3-5 years, ultimately providing recommendations and predictions for the future.

2. Top security concerns

This section will discuss a few of the most prevalent security risks that have been frequently observed in recent times within society.

2.1. Scam. Scams are prevalent and can manifest in various forms, such as emails, text messages, phone calls, and website advertisements. The most prevalent scams include spam emails and long con scams

2.1.1. Spam email. Email is a widely used communication tool, but it is susceptible to receiving spam messages that contain invalid information which can be potentially harmful. These phishing emails have the potential to automatically download harmful software upon opening, expose inappropriate content to minors, or redirect users to fraudulent websites disguised as real ones. These unlawful websites illegitimately collect and retain users' data, which will be used for further phishing purposes. If these pornographic emails are received by children, they can have a significant impact on them. Email remains an invaluable technology despite its issues.

2.1.2. Long con scam. Long con scam prevention possesses a property of a persistent and insidious nature. Fraudsters frequently exploit sympathetic victims, gradually and surreptitiously deceiving them of their money. While phishing is a rapid but transient annoyance, a long con scam can persist for weeks or even months before the victim realises, they have been duped. The similarity among these types of scams is that they become or pretend to be familiar websites or individuals to the victims.

Romance scammer is one of the frequently used identities in long con scams [2]. The perpetrators will pretend to be people of different ages and genders in order to carry out their fraudulent activities. These romance scammers scour the internet and multiple applications searching for their target. They gradually raise their importance and establish a stable relationship with the victims by continuously commenting and interacting with them. Once the relationship is established, the perpetrators begin asking their victims for money, using a variety of lies and excuses.

2.2. DDos/Dos attack. These attacks can destabilise a server. Dos attacks send lots of IP requests to a server. The server will respond to these requests, and because of the TCP Protocol, the server will wait for a response from the sender. But the IP attacks are fake, they would never respond. The server keeps waiting until the time is exceeded. However, the server will receive a new fake IP request. This is how the server ran out of resources. Distributed Denial Service (DDos) attack is based on the Dos attack but a wider attack version. Attackers will attack and control some computers first. The computers will control more computers. All of these computers could be the attacker on the server.

There is constantly a huge concern about networking attacks. The hackers keep implementing new types of DDos attacks, and they can generally be classified into three groups.

Volumetric DDoS Attacks flood the network or service with a massive amount of seemingly legitimate traffic, causing congestion.

TCP State-Exhaustion (Protocol) attacks target the connection state tables in infrastructure components such as load-balancers and firewalls. They could take down high-capacity devices with SYN attacks exploiting layer 3 and layer 4 vulnerabilities.

Application (Resources) Layer DDoS attacks target specific aspects of Layer-7 applications or services and can be difficult to detect.

DOI: 10.54254/2755-2721/47/20241480

Table 1. Comparison of some common attacks

User Datagram Protocol (UDP) flood	Ping (ICMP) flood	Half-Open attack (SYN flood)	HTTP Flood
Volumetric	Volumetric	TCP State-Exhaustion	Application
Overwhelm the target system with lots of data packets.	Occupy the target's inbound bandwidth by sending thousands of ICMP echo requests.	Occupy the server by sending lots of SYN packages and never responding to the server during the TCP three-hand-shake process.	The server is unable to cope with the constant and substantial influx of GET or POST requests, resulting in a malfunction of the website or application.

Table 1 compares the most common types of attack, including UDP, ICMP, SYN and HTTP, highlighting their different characteristics and effects.

Nowadays, DDoS attacks are more frequently occurring in multi-vector form, as they are more challenging to defend. Among the listed attacks, ICMP flood is the one that comes more often with other attacks such as SYN flood.

3. Applications of AI on security

This section explores how AI is being used to address many security challenges, such as detecting spam, preventing DDoS attacks and improving SIEM/IDS systems. These examples highlight the adaptability and potential of AI in enhancing security measures and protecting against emerging threats in our interconnected world.

3.1. Spam email detection. Artificial intelligence is a broad field, and over time numerous researchers have sought to enhance the precision of spam filtering through the use of techniques from various AI subfields.

The concept of employing the Bayesian chain rule for automated spam filtering was first presented in the 20th century [3], and it has generally been successful. Nonetheless, AI can carry the possibility of mistakenly classifying an authentic email as spam [4]. In 2000, [4] highlighted the employment of supervised learning to educate AI models based on Bayesian or memory-based frameworks for spam detection. The findings prove the possibility of this approach. The rate of success with filtering is not always 100%, and the consequences of misclassification are ambiguous. Further research is required to improve the success rate closer to 100%. In a recent study [5], that examined the efficacy of AI models in detecting spam, three different algorithms were tested, all of which achieved over 95% accuracy. However, since fraudsters are capable of coming up with new tactics when previous ones become obsolete, it is crucial that AI be able to autonomously differentiate between spam and legitimate emails.

3.1.1. DDos prevention. AI-powered analytics can effectively handle a considerable number of DDoS attacks originating from a large number of unique IP addresses. Not all IP addresses are authentic. Some are generated by an IP generator tool and therefore have distinctive characteristics. Thus, their behaviours may differ from those of typical users. AI can analyse this data and automatically intercept attacks. As with filtering spam email, intercepting such attacks requires classification.

DOI: 10.54254/2755-2721/47/20241480

Table 2. Data of algorithm accuracies from [6]

Name of Algorithms	Accuracy (%)	Name of Algorithms	Accuracy (%)
J48	100	Random forest	99.99374...
Stochastic Gradient Descent	100	Support Vector Machine	99.99374...
Hybrid	100	Zero -R	99.91236...
Graussian NB Classifier	99.99791...	AR	99.91236...
K nearest Neighbour Classifier	99.99583...		

Multiple AI-based algorithms are available for classification. In a recent study [6], as shown in Table 2, the authors identified that among these algorithms, Stochastic Gradient Decent and J48 are the most accurate algorithms for classifying DDos attacks. The authors have also proposed a hybrid algorithm that has been implemented from parameters of RF, SGD, SVM,K-Nearest Neighbour and Nave Byesare. It stands out with very high accuracy and shows that combining algorithms is a good research direction.

In [7], the journal presents a learning-based detection approach that leverages learning algorithms and stateful/stateless features from Openflow packets to accurately identify low-rate DDoS attacks in SDN control and data planes, demonstrating effective detection with minimal system performance impact.

3.1.2. SIEM/IDS systems. These detection systems detect unusual behaviour on the network. While an Intrusion Detection System (IDS) only detects and records activities, Security Information and Event Management (SIEM) also connects these events together, making it easier for the Security Operations Centre (SOC) to manage network activity. The employment of AI has enhanced the accuracy of identifying malicious cyberattacks, resulting in increased efficiency and precision when using these tools within the SOC.

A significant proportion of existing IDS rely predominantly on historical data. They lack the ability to make predictive inferences. Improving the effectiveness of these systems has the potential to enhance security measures. In [8], the researchers present a new IDS system, called 'Early-A', based on DL principles. Early-A uses a convolutional neural network (CNN) model trained by acquiring knowledge about the characteristics of both normal and malicious network traffic, as extracted from the CICIDS2017 and USTC-TFC-2016 datasets. A predefined threshold is used to classify network traffic as malicious. Deviations from a standard distribution beyond this threshold indicate malicious activity. The concept of building real-time IDS systems has significant benefits in ensuring that these systems remain adaptable and agile when faced with new forms of cyber-attack.

The DL-based AI model proposed by [9] for IDS performs better than conventional ones. The ANOVA F-value based method, impurity-based feature selection, and mutual information-based techniques were used to identify features of the two collected datasets. The AI models were trained using these two datasets and relied on three ML algorithms, K-NN, Decision Trees and Logistic Regression. The accuracy of the decision tree was highest for both datasets, reaching 88% and 98% respectively.

Good AI techniques can make model training progress more effective. For example, [10] compared the performance of two AI-based IDS methods for analysing detected data. The methods are support vector machine (SVM) and deep neural network (DNN) respectively. It turns out that based on the NSL-KDD training dataset, DNN stands out with higher accuracy.

4. Analysis of Challenges

AI is a powerful and efficient technology that can act as a strong defender against web attacks, but it can also become a weapon in the hands of hackers. The inevitability of technology being exploited by hackers requires a discussion of the challenges that AI faces in its development. This section aims to gather and analyze the challenges encountered in the evolution of AI technology and summarise existing solutions and potential strategies.

DOI: 10.54254/2755-2721/47/20241480

4.1. Reliability. The performance of AI depends heavily on extensive data training, which is not always reliable. Two distinct things to humans may appear identical to an AI. In such situations, AI might not always be able to distinguish between them, potentially leading to inaccurate judgments. Furthermore, their decision-making capabilities can be further weakened by a data leakage attack[4].

4.2. Unidentified threat. AI models are trained on existing data, and as a result, these AI models may encounter difficulty in recognising new attacks that are absent from their database In [11], the authors tackle this issue by combining semi-supervised methods to label unfamiliar data and implementing a retraining technique based on sliding windows in their proposed approach. The assessment carried out on the CIC-IDS 2017 dataset, which included a number of DDoS attacks, confirms the effectiveness of this approach in enhancing detection accuracy and ensuring the quality of generated pseudo labels.

4.3. Data Privacy. AI requires access to users' personal data to effectively predict potential attacks or filter spam. However, this creates an invasion of privacy, which is known as the privacy paradox[12]. This issue is unavoidable, as these technologies rely on analysing the data provided by the user. Indeed, every human action on the web is traceable, resulting in a significant increase in the total amount of data on the web. Instead of dwelling on the challenges that come with AI implementation, it is worthwhile to explore how AI can protect data that has not yet been leaked.

4.4. Ethics and morality. AI may appear to mimic human behaviour as it progresses, but it is certainly not subject to human laws. If a criminal uses an AI model or technology to commit a crime, the technique cannot be imprisoned for a long time, as human criminals would be. Similarly, the technology should not be banned on the basis of the actions of a single individual. Consequently, AI will need to develop its own legal framework as society progresses[13]. These regulations should require AI to use data in an ethical way, according to human values and standards.

4.5. Explainable AI. AI training process often operates as a black box with limited transparency into its inner workings using the input dataset. In order to be reliable, AI must be able to explain its logic and that the strategic decisions are visible to humans. The significance of reliable AI is detailed in [14].

The authors described two main approaching techniques: transparent design and post-hoc explainability. Transparent design aims to develop AI models with inherent transparency, making their operations more understandable. Post-hoc explainability methods involve creating simplified surrogate models that imitate the behaviour of complex black-box models.

5. Conclusion

This paper examines the practical applications of AI in addressing modern security challenges and emphasises its importance in security development. The research mainly focuses on detecting spam emails, and preventing DDoS or Dos attacks, as well as the implementation of SIEM and IDS. Analysis indicates that AI provides a pathway for enhancing security detection accuracy and response speed.

From analysed insights in this paper, several crucial deductions emerge. Firstly, AI has the potential to significantly enhance the efficiency of security measures. For instance, the data analysis abilities of AI can be used for the identification and categorisation of harmful packets or messages. It is expected that the development of more efficient AI algorithms is needed to detect and identify new attacks in real time. Furthermore, while AI presents significant potential, it is crucial to ensure its technical reliability and capability to identify emerging threats. ensuring data privacy and complying with legal regulations are equally crucial from ethical and legal perspectives. Additionally, to establish trust within the application of AI, the matter of explainable AI must be tackled. As other concerns are addressed and technology progresses, it is imperative to focus on improving regulations related to AI technology. This will prevent the possible exploitation of advanced technology for illicit purposes. For example, it could prevent the utilization of efficient algorithms to carry out more potent attacks.

DOI: 10.54254/2755-2721/47/20241480

The significance of AI technology within the security domain will become increasingly important in the coming years. With the ongoing progress in technology, the partnership between artificial intelligence and cybersecurity is set to become a vital defence against emerging threats.

In conclusion, this paper analyses the practical progression of AI in the domain of cybersecurity. The significant potential of AI in protecting online security is highlighted, and it will remain a vital area of focus for future research.

References

[1] Z. Zhang, H. A. Hamadi, E. Damiani, C. Y. Yeun and F. Taher, "Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research," in IEEE Access, vol. 10, pp. 93104-93139, 2022, doi: 10.1109/ACCESS.2022.3204051.

[2] Emma Fletcher, 'Romance Scammers' Favorite Lies Exposed', Federal Trade Commission, 2023 <https://www.ftc.gov/news-events/data-visualizations/data-spotlight/2023/02/romance-scammers-favorite-lies-exposed> [accessed 12 October 2023]

[3] I. Androutsopoulos, G. Paliouras, V. Karkaletsis, G. Sakkis, C. D. Spyropoulos and P. Stamatopoulos. "Learning to Filter Spam E-Mail: A Comparison of a Naive Bayesian and a Memory-Based Approach.",2000, https://doi.org/10.48550/arXiv.cs/0009009

[4] X. Yang, W. Luo, L. Zhang, Z. Chen and J. Wang, "Data Leakage Attack via Backdoor Misclassification Triggers of Deep Learning Models," 2022 4th International Conference on Data Intelligence and Security (ICDIS), Shenzhen, China, 2022, pp. 61-66, doi: 10.1109/ICDIS55630.2022.00017.

[5] G. V. Sesha Sai Krishna Vineeth, M. Leela Venkata Sai, M. U. Mahesh, M. Varun and S. Shanmugapriya, "Email Spam: A New Strategy of Screening Spam Emails using Natural Language Processing," 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2023, pp. 710-715, doi: 10.1109/ICAIS56108.2023.10073758.

[6] R. S. Devi, R. Bharathi and P. K. Kumar, "Investigation on Efficient Machine Learning Algorithm for DDoS Attack Detection," 2023 International Conference on Computer, Electrical & Communication Engineering (ICCECE), Kolkata, India, 2023, pp. 1-5, doi: 10.1109/ICCECE51049.2023.10085248.

[7] Haosu Cheng, Jianwei Liu, Tongge Xu, Bohan Ren, Jian Mao, and Wei Zhang. 2020. Machine learning based low-rate DDoS attack detection for SDN enabled IoT networks. Int. J. Sen. Netw. 34, 1 (2020), 56–69. https://doi.org/10.1504/ijsnet.2020.109720

[8] T. Ahmad and D. Truscan, "Efficient Early Anomaly Detection of Network Security Attacks Using Deep Learning," 2023 IEEE International Conference on Cyber Security and Resilience (CSR), Venice, Italy, 2023, pp. 154-159, doi: 10.1109/CSR57506.2023.10224923.

[9] A. Lakshmanarao, A. Srisaila and T. S. Ravi Kiran, "Machine Learning and Deep Learning framework with Feature Selection for Intrusion Detection," 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, 2022, pp. 1-5, doi: 10.1109/IC3IOT53935.2022.9767727.

[10] N. D. Patel, B. M. Mehtre and R. Wankar, "Detection of Intrusions using Support Vector Machines and Deep Neural Networks," 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2022, pp. 1-5, doi: 10.1109/ICRITO56286.2022.9964756.

[11] B. Nugraha, K. Yadav, P. Patil and T. Bauschert, "Improving the Detection of Unknown DDoS Attacks through Continual Learning," 2023 IEEE International Conference on Cyber Security and Resilience (CSR), Venice, Italy, 2023, pp. 477-483, doi: 10.1109/CSR57506.2023.10224989.

[12] G. Smith, "Artificial Intelligence and the Privacy Paradox of Opportunity, Big Data and The Digital Universe," 2019 International Conference on Computational Intelligence and

DOI: 10.54254/2755-2721/47/20241480

Knowledge Economy (ICCIKE), Dubai, United Arab Emirates, 2019, pp. 150-153, doi: 10.1109/ICCIKE47802.2019.9004264.

[13] H. Sun, "Legal Examination and Regulation of Artificial Intelligence Algorithm," 2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI), Kunming, China, 2021, pp. 588-591, doi: 10.1109/CISAI54367.2021.00119.

[14] V. Chamola, V. Hassija, A. R. Sulthana, D. Ghosh, D. Dhingra and B. Sikdar, "A Review of Trustworthy and Explainable Artificial Intelligence (XAI)," in IEEE Access, vol. 11, pp. 78994-79015, 2023, doi: 10.1109/ACCESS.2023.3294569.