EasyChair Preprint

№ 14275

Artificial Intelligence in Cybersecurity

Ghadeer Zidan Suleiman

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

December 16, 2024

Artificial Intelligence In Cybersecurity

Ghadeer Zidan Suleiman
Prince Hussein bin Abdullah College of Information Technology
Irbed, Jordan
ghadeersuliman96@gmail.com

Abstract: This review paper examines the role of artificial intelligence (AI) in advancing cybersecurity. The text explores several AI methodologies, including machine learning, deep learning, and expert systems, and their applications in detecting risks, preventing disruptions, and ensuring data security. The analysis emphasizes the advantages of AI-driven cybersecurity solutions, such as the ability to efficiently process large amounts of data and identify patterns that indicate possible security weaknesses. While ethical and privacy problems are addressed, the usefulness of AI in detecting malware, network breaches, and spam is emphasized. Regardless, the material also discusses the applications, limitations, and ethical considerations associated with the use of AI in cybersecurity, highlighting the need for a balanced strategy that integrates technological advancement with human expertise and supervision.

Keywords: Cybersecurity, Artificial Intelligence (AI), Machine Learning (ML), Expert system (ES), Deep learning (DL), IOT, HLCSM, NIST, Explanation Artificial Intelligence (XAI)

I. INTRODUCTION

Artificial intelligence (AI) is a domain of computer science dedicated to the creation of intelligent agents; to do this, robots must undergo accurate learning, necessitating training using learning algorithms. AI methodologies depend on algorithms but may also utilize extensive data and substantial computational power to learn by brute force. AI operates in three modalities: aided intelligence, enhanced intelligence, and autonomous intelligence, which are systems proficient in independent thinking, learning, and decision-making. Artificial intelligence has several uses across various disciplines, including healthcare, finance, manufacturing, and, more recently, cybersecurity. [8,9].

In cybersecurity, Artificial Intelligence (AI) is a compelling technology that may offer advanced analysis and insights to combat always evolving threats. It achieves this by swiftly assessing extensive datasets and surveilling various forms of cyber threats. Technology is being integrated into cybersecurity to automate security operations or support human security teams. [8,9,10].

Cybersecurity involves the use of many tactics, techniques, and resources to protect systems from possible threats and vulnerabilities, while effectively providing precise services to users. [2].

Cybersecurity seeks to minimize threats to the maximum degree possible while swiftly and effectively addressing the requirements for detection, response, and recovery from events. [2].

An expert system (ES), usually referred to as a knowledge-based system, comprises an information repository and an inference engine that facilitates logical thinking and problem-solving. Their problem-solving abilities comprise two distinct methodologies: case-based reasoning, which involves leveraging past difficulties and their answers for new challenges, and rule-based reasoning, which relies on expert-defined criteria to address issues. Case-based reasoning involves evaluating previous circumstances and adapting answers appropriately, whereas rule-based reasoning use rules that consist of a condition and a corresponding action. Rule-based systems are incapable of autonomously acquiring new rules or modifying existing ones. unlike case-based systems. Expert systems (ESs) can be utilized to provide decision-making assistance in cyberspace by examining altered data from security systems to identify the presence of malicious network or system activity. They have the ability to conduct real-time monitoring in digital settings and provide alarm alerts and pertinent information for security experts to implement suitable steps.[11].

Machine learning (ML) encompasses a set of methodologies that allow computers to gain information and enhance their performance autonomously, without requiring explicit directives or programming. It aids in the discovery and formalization of data principles inside systems, fosters learning from data, and improves performance via experience. Machine learning employs statistical techniques to extract insights, identify patterns, and draw conclusions. It may be categorized into three main types: supervised learning, unsupervised learning, and reinforcement learning. Common machine learning techniques utilized in cybersecurity encompass the decision trees, support vector machines, Bayesian approaches, and ensemble learning. [12,13].

The real-time examination of large data sets using machine learning algorithms facilitates the detection of possible security issues. Collaboration, technological advancement, and user awareness are critical elements for effective cybersecurity. Nevertheless, the use of AI and ML technologies offers both progress and new Concerns, as unscrupulous individuals exploit them to conduct attacks and carry out phishing schemes. Ensuring the ethical application of AI in cybersecurity is crucial to prevent its exploitation. Artificial intelligence (AI) aids researchers in understanding the intricacies of ecological processes and provides essential insights for conservation initiatives. Artificial intelligence (AI) integrated transportation systems improve route optimization, reduce emissions, and increase operational efficiency [3].

The growing prevalence of machine learning (ML) drives research into algorithms that elucidate ML models and their predictions, referred to as eXplainable Artificial Intelligence (XAI) [20].

- XAI Frameworks

XAI frameworks are instruments that provide reports detailing a model's functionality and endeavor to elucidate its operational mechanisms. Notable XAI frameworks comprise SHAP, LIME, ELI5, Skater, DALEX, and ALE. SHAP is a paradigm for elucidating and justifying the outcomes of predictive models, use game theory to illustrate the correlation between optimum credit allocation and localized explanations [42]. LIME is analogous to SHAP but functions more rapidly, offering elucidations for the influence of each feature in a data sample [43]. ELI5 is an explainability package developed by MIT that enhances machine learning and allows for direct comparison of models across different frameworks and packages [44]. Skater is a model-agnostic framework for model interpretation across many models, whereas DALEX aids researchers in understanding model behavior [45]. ALE is a global explanation technique that examines the relationship between feature values and target variables, demonstrating the fundamental effects of individual predictors and their second-order interactions in opaque supervised learning models [46]. These frameworks aim to improve the transparency and accessibility of the opaque nature of machine learning to humans.

Deep learning: also known as deep neural learning, utilizes data to train computers to accomplish tasks that humans can do. Deep learning algorithms mimic the cognitive processes of the human brain to evaluate data and produce patterns that influence decision-making. They have the ability to carry out iterative tasks, making alterations to the job to enhance the outcomes. Cybersecurity utilizes deep learning techniques to handle the vast volumes of data that are gathered daily. Deep learning methods enable the implementation of supervised, unsupervised, and reinforcement learning approaches. Xu et al. conducted a case study to assess the efficacy of deep learning in identifying network intrusions. This showcases the capabilities of AI-driven technologies to do instantaneous analysis and precisely detect harmful network traffic [5,13].

IOT: The term "Internet of Things" encompasses the interconnected network of devices and the technology that enables communication between these devices and the cloud, as well as between the devices themselves. An essential aspect of the future of cybersecurity centers on the Internet of Things (IoT). The Internet of Things (IoT) consists of an extensive network of networked objects, ranging from intelligent appliances and wearable gadgets to industrial systems and essential infrastructure. Although the Internet of Things (IoT) offers unparalleled ease and automation, it also brings about weaknesses that may be easily exploited by malevolent individuals. Inadequate security measures, weak authentication mechanisms, and subpar device management can render IoT systems susceptible to attacks. To counter these risks, future cybersecurity strategies must prioritize robust encryption protocols, regular software updates, and enhanced security measures tailored specifically for IoT devices [1].

A. The Human-in-the-Loop Cyber Security Model

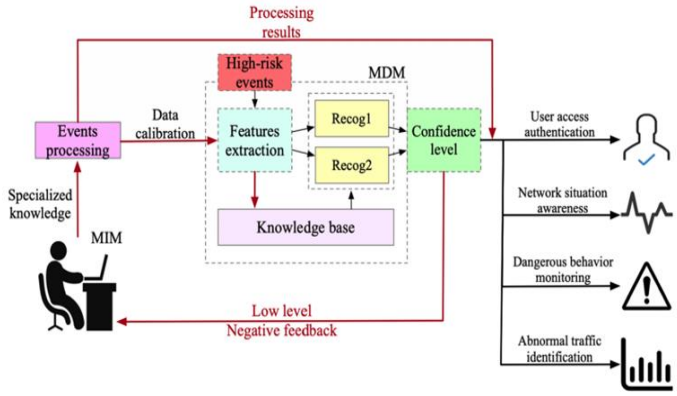

Figure 1: Human-in-the-Loop Cyber Security Model (HLCSM)

Figure 1 shows The Human-in-the-Loop Cyber Security Model (HLCSM) is a novel approach that seeks to combine human experience with machine intelligence in the realm of cyber security. Artificial intelligence (AI) technology provides significant advantages in several applications, despite its inherent constraints. The model is divided into two subordinate modules: the Machine Detection Module (MDM) and the Manual Intervention Module (MIM). The main role of MDM is to proactively avoid and detect cyber issues, while also ensuring data readiness and extracting pertinent attributes. MIM functions as a supplementary entity, overseeing events by the application of experienced knowledge. The Confidence Level Module (CLM) is designed to build a smooth link between MDM and MIM, enabling efficient collaboration. The CLM combines the results and determines the Confidence Level, therefore maximizing human resources and reducing the time needed for identification. However, when the Confidence Level is low, experts carefully examine the information to minimize the likelihood of errors. The primary goal of the HLCSM is to augment the proficiency and reliability of cyber security systems by combining human knowledge with machine intelligence. However, it is essential to recognize that the best results can only be achieved by integrating AI with human-in-the-loop technology [2].

B. NIST : The National Institute of Standards and Technology

Figure 2: NIST cybersecurity framework.

Figure 2 shows The National Institute of Standards and Technology (NIST) is a voluntary framework that aims to assist enterprises in comprehending, controlling, and mitigating cybersecurity risks [41]. The framework comprises four components: Functions, Categories, Subcategories, and Informative references. The initial two tiers of the framework, comprising of 5 cybersecurity functions and 23 solution categories, offer a complete perspective on cybersecurity management. The suggested taxonomy incorporates an additional tier that delineates AI-driven applications that align with each level of the framework. This taxonomy offers a precise and straightforward classification of current research on AI for the field of cybersecurity, making it easy to understand and navigate [1].
The Gordon-Loeb (GL) Model is presented to assist businesses in incorporating cost-benefit analysis into the NIST Cybersecurity Framework. This model posits that companies are susceptible to cybersecurity breaches, with the likelihood of a violation equating to the level of vulnerability. The ideal degree of cybersecurity investment is established by reducing the aggregate projected costs of security breaches with the investment expenditure [15].

II. RESEARCH METHODOLOGY

This review paper's methodology involves conducting a comprehensive analysis of contemporary literature published between 2020 and 2024. The research will include an analysis of the benefits, limitations, strengths, and risks of AI-based cybersecurity approaches. explores the use of artificial intelligence (AI) in cybersecurity, specifically in the areas of user access authentication, network state knowledge, hostile activity monitoring, and anomalous traffic recognition. This paper examines the application of artificial intelligence (AI) in intrusion detection systems (IDS) and examines the ethical consequences linked with it. The research categorizes AI approaches such as machine learning and deep learning, along with their applications in threat detection, intrusion detection systems (IDS), and real-time data processing.

What is the crucial function of Artificial Intelligence (AI) in cybersecurity, specifically looking at how it is used in areas such as identifying threats, assessing vulnerabilities, responding to incidents, and doing predictive analyses [3]?
What is the paradoxical nature of using AI in cybersecurity, where AI can be used both for public good and for harm [4]?

What is the potential of Artificial Intelligence (AI), specifically sophisticated language models such as ChatGPT, in improving the capacity of Intrusion Detection Systems (IDS) to recognize, categorize, and detect abnormal network traffic and cyber-attacks [5]?

What is the significance of ML in the field of cybersecurity, with a special emphasis on the identification of threats and the implementation of protective measures [6]?
What are the impacts and limitations of artificial intelligence in cybersecurity [7]?

III. LITERATURE REVIEW

The study by Katanosh Morovat and Brajendra Panda,2020 explains that the growing complexity of cyberattacks has required the creation of sophisticated cybersecurity methods. AI technologies have been employed to protect systems from a range of threats, including effective defensive capabilities to identify and respond to malware attacks, network intrusions, phishing and spam emails, and data breaches. AI techniques, including learning algorithms, expert systems, machine learning, deep learning, and biologically inspired computation, are crucial topics in the field of cybersecurity. Artificial Intelligence (AI) can effectively analyze vast quantities of data and derive insights from previous security breaches to anticipate forthcoming cyber threats. Nevertheless, artificial intelligence (AI) is constrained by factors such as the need for extensive data, frequent occurrence of false alarms, and susceptibility to possible assaults. Scientists have devised techniques to categorize and identify malicious software by utilizing approaches such as data mining and machine learning. Current research has mostly concentrated on using deep learning architectures to identify sophisticated malicious software. Artificial intelligence can greatly enhance data and application security in the future. However, there are ongoing worries over the dependability and potential risks linked with AI.[14]

A study conducted by Nicolas Camacho,2024 Artificial intelligence systems can rapidly evaluate large amounts of data to detect unusual patterns that may indicate possible security breaches. These technologies allow enterprises to take proactive measures to prevent hazards and protect sensitive information. Nevertheless, the utilization of AI in cybersecurity also presents ethical and privacy concerns, requiring a measured approach to its adoption. This study provides a thorough analysis of the advantages, restrictions, and ethical considerations of artificial intelligence (AI) in the field of cybersecurity. It highlights the need of achieving a harmonious equilibrium between technological advancement and ethical obligations. The trajectory of cybersecurity is shaped by the widespread use of digital technology, such as the Internet of Things (IoT), which exposes potential weaknesses that may be exploited by malevolent individuals. Future cybersecurity policies should give top priority to implementing strong encryption methods, ensuring frequent software upgrades, and implementing better security measures designed particularly for Internet of Things (IoT) devices. Effective collaboration among manufacturers, developers, and cybersecurity specialists is crucial to guarantee that IoT devices are developed with security as a primary consideration right from the beginning.[2]

A study conducted by Roba Abbas and colleagues, 2023 highlights the contradictory characteristics of AI in cybersecurity, presenting several challenges, such as its inherent fallibility, its role within a larger socio-technical framework, the potential negative consequences of unregulated AI, and concerns over the accuracy and fairness of data. Understanding the complex socio-technical system and the potential risks associated with AI in cybersecurity is crucial, as mentioned in the conclusion of the research. The statement emphasizes the need for highly skilled

professionals in the areas of cybersecurity and risk management, while also emphasizing the need to maintain a balance between technology, ethics, and regulation. When integrating AI into cybersecurity, it is essential to thoroughly evaluate the characteristics of the data, potential risks, and the probability of unforeseen consequences [3].

This study by Michal Markevich and Maurice Dawson, 2023, examines the potential of Artificial Intelligence (AI) to improve intrusion detection systems (IDS) in the cybersecurity domain. This demonstrates that artificial intelligence (AI) is a crucial asset in enhancing the accuracy of intrusion detection systems (IDS) in recognizing and responding to cyber-attacks. However, the study also highlights the limitations and challenges of integrating artificial intelligence (AI) into intrusion detection systems (IDS), such as the complexity of calculations and the potential for biases in the training data. Deploying sophisticated language models like ChatGPT can enhance cybersecurity measures, but it is crucial to tackle these issues to offer a more robust defense against complex cyber threats. The study indicates that artificial intelligence (AI) can improve the precision of intrusion detection systems (IDS). However, it also faces challenges like as inaccurate positive and negative results, intricate computational demands, limitations in resources, and worries about data privacy [5].

Another research by Ugochukwu Okoli et al,2024 examines the importance of Machine Learning (ML) in cybersecurity, specifically its applications in threat detection and defense systems. The versatility of machine learning enables it to detect nuanced patterns in extensive datasets, rendering it highly valuable in the realm of cyber warfare. The paper highlights the necessity of adopting a holistic strategy that integrates technology with ethical issues, blending human expertise with machine intelligence. Additionally, it explores the difficulties and advantages of ensuring cybersecurity in power grids and maritime industries, as well as the influence of the Internet of Things (IoT) on cyber threats. The report asserts that the integration of machine learning into cybersecurity is essential for organizations to effectively counteract the ever-changing threats [6].

The latest research in this article by Miraj Ansari and colleagues,2022 examines the significant impact of AI on cybersecurity as it enables intelligent systems and robots to mimic human behavior. AI platforms enable the deployment of machine learning and deep learning models in businesses, therefore enhancing data security, reducing reliance on cybersecurity experts, and lowering costs associated with maintenance and auditing. Artificial Intelligence (AI) improves the effectiveness of network intrusion detection, vulnerability management, and data center security. During the ongoing COVID-19 pandemic, there has been a consistent increase in investments in artificial intelligence (AI), which allows for the continuous monitoring of vulnerability databases in real-time. However, artificial intelligence (AI) does have limitations, including the potential for manipulation by unscrupulous persons, the impossibility to completely replace human expertise, and difficulties in adapting to constantly evolving threats. The high costs involved in implementing AI-powered cybersecurity solutions and the possibility of hostile actors reverse engineering AI systems highlight the need for continuous improvements in system security [7].

IV. APPLICATIONS

Artificial Intelligence (AI) has become increasingly used in cyber security applications, including intrusion detection, malware detection, and spam filtering. However, most AI-based techniques are deployed in a "black-box" manner, making it difficult for security experts and customers to explain how they reach conclusions. The absence of openness and interpretability may diminish user confidence in cyber protection models. XAI should be included in cybersecurity models to develop explainable frameworks while preserving high accuracy.

1. SPAM

Spam has emerged as a substantial problem for internet users, constituting approximately 55% of all emails dispatched in 2021[16,17]. AI-based systems constitute an excellent answer to this issue due to their ability for self-evolution and optimization. Nonetheless, the privacy and legal intricacies of spam lead users to scrutinize the efficacy of AI models [18], particularly black-box machine learning and deep learning models [21]. XAI algorithms have been employed to enhance ML models with attributes such as explainability and transparency [20]. Numerous research has investigated the identification of fraudulent spam news using machine learning algorithms, including the SHAP technique and HateXplain. XAI may enhance system trust and mitigate automation bias in botnet detection [19], while the integration of AI methodologies can be beneficial for regulatory compliance and provide real-time explanations for fraud prevention and model accuracy assessment. XAI has been utilized in several fields, including cybersecurity, algorithmic domain generation, and denial-of-service attacks. Nonetheless [20], the privacy and legal complexities of spam lead consumers to doubt the efficacy of AI models [22]. The predominant kinds of harmful spam globally encompass Trojan horses, malware, and ransomware. Numerous strategies have been created to address the issue of spam [23]. Currently, three methods to alleviate such assaults are prominent: Concentrate on awareness, blacklists, and machine learning (ML). Recently, Deep Learning (DL) has emerged as one of the most effective strategies in machine learning [24].

2. FRAUD

Personal account breaches and online financial fraud increased during the Covid-19 pandemic, costing businesses and individuals £130 billion annually and the global economy $3.89 trillion [25]. AI systems can be employed to counteract fraud assaults. However, practical challenges arise in implementing AI techniques, especially in using Explainable Artificial Intelligence (XAI) to make sense of the conclusions and forecasts generated by intricate models [26]. Studies have looked into the rationale behind fraud detection with both supervised and unsupervised models, and some results

suggest that combining the two methods could be beneficial [27]. XAI methodologies can enhance the efficacy of fraud detection models, with certain models attaining overall accuracy and AUC of 94% and 96.9%, respectively. Innovative fraud detection algorithms like Fraud Memory and FinDeepBehaviorCluster perform comparably to the classic HBD-SCAN, however, it displays computational efficiency that is hundreds of times superior [28].

To improve threat detection, prediction, and mitigation, banks are integrating cybersecurity and fraud operations. Developing digital trust, adopting a "customer journey" approach to fraud, and modernizing internal and customer operations are all necessary to achieve this. The unified operating model focuses on people, data, technology, processes, activities, and governance [29].

3. NETWORK INTRUSION

Network intrusions are unauthorized infiltrations into a company's computer or domain. Network Intrusion Detection Systems (NIDS) monitor network activity for unusual behavior. Recent works have adopted ML and DL algorithms for efficient NIDSs. Explainability is being considered to make NISDs more robust. Two-staged pipelines have been proposed for robust NIDS [30].

Zakaria et al. developed a novel DL and XAI-based system for intrusion detection in IoT networks, using three explanation methods: LIME, SHAP, and RuleFit [31]. This system was tested on NSL-KDD and UNSW-NB15 datasets [32], demonstrating its effectiveness in strengthening IoT IDS interpretability against well-known attacks. Yiwen et al. presented an intrusion detection system for malicious traffic intrusion, using XAI-based methods and neural networks [33]. Sivamohan et al. presented BiLSTM-XAI, reducing the complexities of BiLSTM models to enhance detection accuracy and explainability [34]. Hong et al. proposed FAIXID, a network intrusion detection framework using XAI and data cleaning techniques to enhance explainability and understanding of alerts [35]. Basim et al. used the Decision Tree algorithm for trust management and demonstrated its advantages [36]. Syed et al. proposed a three-stage architecture to detect malicious intrusion in network traffic, achieving higher accuracy rates [37].

4. DOMAIN GENERATION ALGORITHMS (DGA)

DGAs are viruses that produce several domain names for covert communication with Command and Control (C2) servers. Because there are so many different domain names, traditional techniques like sink-holing and blacklisting are insufficient. Mitigating DGA tactics poses difficulties, as administrators must identify the virus, DGA, and seed value to exclude hazardous networks and servers [38]. Machine learning classifiers have been proposed to identify domain generation algorithms (DGAs) responsible for generating certain domain names and initiating targeted corrective measures. Nonetheless, evaluating the internal logic is difficult because of the opaque nature. Franziska et al. provided a visual analytics framework for the classification of DGAs; however, this does not inherently indicate the way the model categorizes the data [39]. Arthur et al. introduced two ResNet-based detection classifiers for binary and multiclass classification, demonstrating strong performance in both categories. The explainability research revealed that several self-learned attributes utilized by deep learning systems were also applied in classifiers [40].

V. CONCLUSION

In cybersecurity, artificial intelligence (AI) has shown great potential for enhancing response times, automating detection processes, and fortifying security protocols. Expert systems, deep learning, and machine learning are crucial AI methods for identifying and addressing emerging cyber threats. However, while using AI in this industry, it is imperative to apply responsible and ethical techniques. Protecting sensitive data and vital infrastructure requires a well-thought-out plan that blends human knowledge with AI skills. This study emphasizes how AI may help with important problems including spam filtering, fraud detection, network infiltration, and domain generation algorithm (DGA) resistance. Ultimately, the collaboration between human analysts and AI-driven systems will be vital in navigating the complexities of modern cybersecurity challenges and ensuring robust protection against sophisticated threats.

REFERENCES

[1] kaur, R., Gabrijelčič, D., & Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion , 97* , 101804

[2] Zhang, Z., Ning, H., Shi, F., Farha, F., Xu, Y., Xu, J., ... & Choo, K. K. R. (2022). Artificial intelligence in cyber security: research advances, challenges, and opportunities. *Artificial Intelligence Review*, 1-25.

[3] Camacho, N. G. (2024). The Role of AI in Cybersecurity: Addressing Threats in the Digital Age. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *3*(1), 143-154.

[4] Michael, K., Abbas, R., & Roussos, G. (2023). AI in cybersecurity: The paradox. IEEE Transactions on Technology and Society, 4(2), 104-109.

[5] Markevych, M., & Dawson, M. (2023, July). A review of enhancing intrusion detection systems for cybersecurity using artificial intelligence (ai). In *International conference Knowledge-based Organization* (Vol. 29, No. 3, pp. 30-37).

[6] Okoli, U.I., Obi, O.C., Adewusi, A.O., & Abrahams, T.O. (2024). Machine learning in cybersecurity: A review of threat detection and defense mechanisms. *World Journal of Advanced Research and Reviews , 21* (1), 2286-2295.

[7] Ansari, M. F., Dash, B., Sharma, P., & Yathiraju, N. (2022). The impact and limitations of artificial intelligence in cybersecurity: a literature review. *International Journal of Advanced Research in Computer and Communication Engineering*.

[8] [1] John McCarthy," Artificial Intelligence logic and formalizing common sense," Stanford University, CA, USA

[9] Lidestri, N., Maher, Stephen J., & Zunic, Nev.," The Impact of Artificial Intelligence in Cybersecurity,". ProQuest Dissertations and Theses, 2018.

[10] Russell Stuart J., Norvig, Peter (2003), " Artificial Intelligence: A Modern Approach, ". (3rd ed.), Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13- 790395-2.

[11] Nadine Wirkuttis, Hadas Klein, "Artificial Intelligence in Cybersecurity,". Cyber, Intelligence, and Security, Volume 1, No. 1, January 2017.

[12] Machine Learning Methods for Malware Detection. Kaspersky Lab, 2020.

[13] Thanh Cong Truong, Quoc Bao Diep, Ivan Zelinka, "Artificial Intelligence in the Cyber Domain: Offence and Defense,". Symmetry Journal, March 2020.

[14] Morovat, K., & Panda, B. (2020, December). A survey of artificial intelligence in cybersecurity. In *2020 International conference on computational science and computational intelligence (CSCI)* (pp. 109-115). IEEE.

[15] Gordon, L. A., Loeb, M. P., & Zhou, L. (2020). Integrating cost–benefit analysis into the NIST Cybersecurity Framework via the Gordon–Loeb Model. *Journal of Cybersecurity*, *6*(1), tyaa005.

[16] E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid, A. O. Adetunmbi, and O. E. Ajibuwa, ''Machine learning for email spam filtering: Review, approaches and open research problems,'' Heliyon, vol. 5, no. 6, Jun. 2019, Art. no. e01802, doi: 10.1016/j.heliyon.2019.e01802.

[17] Daily Number of E-Mails Worldwide 2025. Statista. Accessed: Oct. 1, 2024. [Online]. Available: https://www.statista.com/statistics/456500/daily-number-of-e-mails-worldwide/.

[18] A. Karim, S. Azam, B. Shanmugam, K. Kannoorpatti, and M. Alazab, ''A comprehensive survey for intelligent spam email detection,'' IEEE Access, vol. 7, pp. 168261–168295, 2019, doi: 10.1109/ACCESS.2019.2954791.

[19] B. Mathew, P. Saha, S. M. Yimam, C. Biemann, P. Goyal, and A. Mukher2740 jee, ''HateXplain: A benchmark dataset for explainable hate speech 2741 detection,'' in Proc. AAAI Conf. Artif. Intell., May 2021, vol. 35, no. 17, 2742 pp. 14867–14875.

[20] M. Renftle, H. Trittenbach, M. Poznic, and R. Heil, ''Explaining any ML model?—On goals and capabilities of XAI,'' 2022, arXiv:2206.13888. 2727.

[21] R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman, ''Metrics for explainable AI: Challenges and prospects,'' 2018, arXiv:1812.04608.

[22] T. Almeida, J. M. Hidalgo, and T. Silva, ''Towards SMS spam filtering: Results under a new dataset,'' Int. J. Inf. Secur. Sci., vol. 2, no. 1, Art. no. 1, Mar. 2013

[23] Zhao, W., Zhu, Y.: An email classification scheme based on decision-theoretic rough set theory and analysis of email security. In: Proceedings of the TENCON 2005-2005 IEEE Region 10 Conference, pp. 1–6. IEEE (2005)

[24] Vinayakumar, R., Soman, K., Poornachandran, P., Akarsh, S., Elhoseny, M.: Deep learning framework for cyber threat situational awareness based on email and url data analysis. In: Hassanien, A.E., Elhoseny, M. (eds.) Cybersecurity and Secure Information Systems, pp. 87–124. Springer, New York (2019)

[25] J. Gee and P. M. Button, ''The financial cost of fraud 2019,'' Tech. Rep., 2019, p. 28.

[26] I. Psychoula, A. Gutmann, P. Mainali, S. H. Lee, P. Dunphy, and F. Petitcolas, ''Explainable machine learning for fraud detection,'' Computer, vol. 54, no. 10, pp. 49–59, Oct. 2021, doi: 10.1109/MC.2021.3081249.

[27] IEEE-CIS Fraud Detection. Accessed: Oct. 23, 2024. [Online]. Available: https://kaggle.com/competitions/ieee-fraud-detection.

[28] W. Min, W. Liang, H. Yin, Z. Wang, M. Li, and A. Lal, ''Explainable deep behavioral sequence clustering for transaction fraud detection,'' 2021, arXiv:2101.04285.

[29] Hasham, S., Joshi, S., & Mikkelsen, D. (2019). Financial crime and fraud in the age of cybersecurity. *McKinsey & Company*, *2019*.

[30] P. Barnard, N. Marchetti, and L. A. D. Silva, ''Robust network intrusion detection through explainable artificial intelligence (XAI),'' IEEE Netw. Lett., early access, Oct. 23, 2024, doi: 10.1109/LNET.2022.3186589.

[31] Z. A. E. Houda, B. Brik, and L. Khoukhi, '''Why should i trust your IDS?': 2870 An explainable deep learning framework for intrusion detection systems in Internet of Things networks,'' IEEE Open J. Commun. Soc., vol. 3, pp. 1164–1176, 2022, doi: 10.1109/OJCOMS.2022.3188750.

[32] N. Moustafa and J. Slay, ''UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set),'' in Proc. Mil. Commun. Inf. Syst. Conf. (MilCIS), Nov. 2015, pp. 1–6, doi: 2527 10.1109/MilCIS.2015.7348942.

[33] (Oct. 21, 2024). Network Intrusion Detection Based on Explainable Artificial Intelligence. Accessed: Oct. 2, 2022. [Online]. Available: https://www.researchsquare.com

[34] (Oct. 23, 2024). KHO-XAI: Krill Herd Optimization and Explainable Artificial Intelligence framework for Network Intrusion Detection Systems in Industry 4.0. Accessed: Oct. 23, 2024. [Online]. Available: https://www.researchsquare.com

[35] H. Liu, C. Zhong, A. Alnusair, and S. R. Islam, ''FAIXID: A framework for enhancing AI explainability of intrusion detection results using data cleaning techniques,'' J. Netw. Syst. Manage., vol. 29, no. 4, p. 40, May 2021, doi: 10.1007/s10922-021-09606-8.

[36] B. Mahbooba, M. Timilsina, R. Sahal, and M. Serrano, ''Explainable artificial intelligence (XAI) to enhance trust management in intrusion detection systems using decision tree model,'' Complexity, vol. 2021, pp. 1–11,Oct. 2024, doi: 10.1155/2021/6634811

[37] S. Wali and I. Khan, ''Explainable AI and random forest based reliable intrusion detection system,'' TechRxiv, Dec. 2021, doi: 10.36227/techrxiv.17169080.v1.

[38] Y. Li, K. Xiong, T. Chin, and C. Hu, ''A machine learning framework for domain generation algorithm-based malware detection,'' IEEE Access, vol. 7, pp. 32765–32782, 2019, doi: 10.1109/ACCESS.2019.2891588.

[39] F. Becker, A. Drichel, C. Müller, and T. Ertl, ‘‘Interpretable visualizations of deep neural networks for domain generation algorithm detection,’’ in Proc. IEEE Symp. Vis. Cyber Secur. (VizSec), Oct. 2020, pp. 25–29, doi: 10.1109/VizSec51108.2020.00010.

[40] A. Drichel, N. Faerber, and U. Meyer, ‘‘First step towards EXPLAINable DGA multiclass classification,’’ in Proc. 16th Int. Conf., Rel. Secur., New York, NY, USA, Aug. 2021, pp. 1–13, doi: 10.1145/3465481.3465749.

[41] B. Pranggono and A. Arabo, “Covid-19 pandemic cybersecurity issues,” Internet Technology Letters, vol. 4, no. 2, p. e247, 2021

[42] Z. Pan, S. Fang, and H. Wang, “Lightgbm technique and differential evolution algorithm-based multi-objective optimization design of ds-apmm,” IEEE Transactions on Energy Conversion, vol. 36, no. 1, pp. 441–455, 2021.

[43] M. S. Kamal, A. Northcote, L. Chowdhury, N. Dey, R. G. Crespo, and E. Herrera-Viedma, “Alzheimerâ€™s patient analysis using image and gene expression data and explainable-ai to present associated genes,” IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–7, 2021

[44] M. Boldt, S. Ickin, A. Borg, V. Kulyk, and J. Gustafsson, “Alarm prediction in cellular base stations using data-driven methods,” IEEE Transactions on Network and Service Management, vol. 18, no. 2, pp. 1925–1933, 2021.

[45] V. Arya, R. K. Bellamy, P.-Y. Chen, A. Dhurandhar, M. Hind, S. C. Hoffman, S. Houde, Q. V. Liao, R. Luss, A. Mojsilović et al., “One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques,” arXiv preprint arXiv:1909.03012, 2019.

[46] P. Biecek, “Dalex: explainers for complex predictive models in r,” The Journal of Machine Learning Research, vol. 19, no. 1, pp. 3245–3249, 2018.