Unveiling Shadows: Harnessing Artificial Intelligence for Insider Threat Detection

Erhan Yilmaz
Department of Computer Engineering, Ege University, Türkiye | Kron Technologies, Türkiye
erhan.yilmaz@itu.edu.tr (corresponding author)

Ozgu Can
Department of Computer Engineering, Ege University, Türkiye
ozgu.can@ege.edu.tr

Received: 15 January 2024 | Revised: 1 February 2024 | Accepted: 4 February 2024

 DOI: https://doi.org/10.48084/etasr.6911

ABSTRACT

Insider threats pose a significant risk to organizations, necessitating robust detection mechanisms to safeguard against potential damage. Traditional methods struggle to detect insider threats operating within authorized access. Therefore, the use of Artificial Intelligence (AI) techniques is essential. This study aimed to provide valuable insights for insider threat research by synthesizing advanced AI methodologies that offer promising avenues to enhance organizational cybersecurity defenses. For this purpose, this paper explores the intersection of AI and insider threat detection by acknowledging organizations' challenges in identifying and preventing malicious activities by insiders. In this context, the limitations of traditional methods are recognized, and AI techniques, including user behavior analytics, Natural Language Processing (NLP), Large Language Models (LLMs), and Graph-based approaches, are investigated as potential solutions to provide more effective detection mechanisms. For this purpose, this paper addresses challenges such as the scarcity of insider threat datasets, privacy concerns, and the evolving nature of employee behavior. This study contributes to the field by investigating the feasibility of AI techniques to detect insider threats and presents feasible approaches to strengthening organizational cybersecurity defenses against them. In addition, the paper outlines future research directions in the field by focusing on the importance of multimodal data analysis, human-centric approaches, privacy-preserving techniques, and explainable AI.

Keywords-cybersecurity; insider threats; artificial intelligence

I. INTRODUCTION

Insider threats present a considerable security risk, as malicious actors within an organization can heavily damage it through unauthorized access to sensitive information. Insider threats include current or former employees, contractors, and business partners who leverage their sanctioned access privileges to an organization's networks, systems, or data to intentionally compromise the confidentiality, integrity, or availability of the organization's information technology assets or the information itself. The gravity of the insider threat stems from the insider's knowledge of internal processes and systems, and their ability to intentionally misuse their authorized access to negatively affect the organization through data theft, sabotage, or fraud. Developing robust insider threat programs to detect and prevent such risks remains an imperative but challenging endeavor for most enterprises and institutions [1]. In [2], it was discovered that 27% of cybercrime incidents were suspected to be carried out by individuals within the organization, and 30% of those surveyed believed that insiders caused greater harm compared to external attackers. In [3], internal fraudsters were identified as the main culprits in 29% of economic crime cases.

Detecting insider threats is challenging, as malicious insiders can operate within their authorized access permissions, and their actions may appear harmless when examined in isolation. Artificial Intelligence (AI) techniques offer potential solutions for more effective insider threat detection. AI can identify patterns and anomalies in user behavior that may indicate malicious intent before serious damage occurs [4]. An efficient insider threat detection program is imperative given the increasing number and severity of insider incidents. This study aims to address this primary concern. Even with AI's potential, most enterprises and institutions still find it difficult to put solid solutions into practice. This paper presents an overview of the current AI techniques utilized to detect insider threats, discusses the major challenges, and highlights potential research and development opportunities in the field to provide a comprehensive understanding. In addition, it explores cutting-edge approaches involving machine learning, natural language processing, knowledge graphs, and adversarial AI to identify insider threats. Furthermore, it addresses the main

challenges related to model interpretability, bias reduction, and privacy protection.

II. INSIDER THREAT

The insider threat describes the risk that someone who has authorized access to sensitive information or systems within an organization may misuse that access to compromise or harm it. This includes compromising the integrity, confidentiality, or availability of the organization's data, personnel, or physical premises. Since the threat comes from an insider, they already have some level of trust and access, making this type of threat especially risky [5]. Insider threats originate from individuals within the organization, such as regular staff, freelancers, interns, and other personnel linked to the company. These internal actors possess different degrees of trust and authority [6]. The total average cost of insider threat incidents was estimated to be $15.4 million in 2022, up 37% over previous years [7]. The patterns shown in Figure 1 highlight the necessity of implementing robust insider threat mitigation programs.

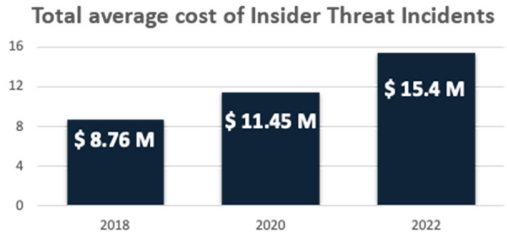

Fig. 1. Total average cost of insider threat incidents.

The Carnegie Mellon University-based CERT division has identified four distinct categories for malicious insider activity [8], which are shown in Figure 2:

- Information Technology (IT) Sabotage: This involves insiders employing IT resources to intentionally cause harm to either an organization or an individual.
- Intellectual Property (IP) Theft: Insiders exploit IT tools to steal intellectual property from the organization. This category encompasses instances of industrial espionage, which might also involve external actors.

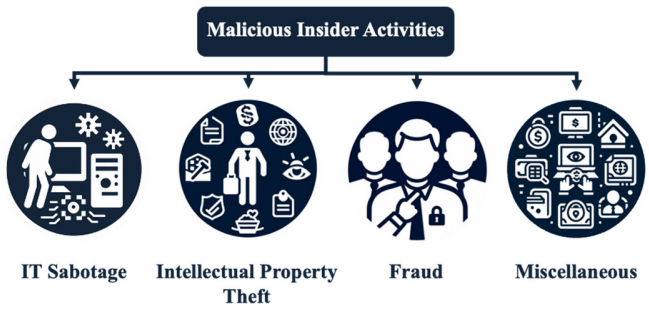

Fig. 2. Classification of malicious insider activities.

- Fraud: Insiders misuse IT capabilities to illicitly alter, append, or remove an organization's data (excluding programs or systems) with the intention of personal benefit. Additionally, this class encompasses instances where information theft leads to crimes such as identity theft or credit card fraud.
- Miscellaneous: This category encompasses cases where an insider's actions do not align with IP theft, fraud, or IT sabotage objectives.

The threat of insider security breaches presents a significant business vulnerability stemming from inadequate breach detection, delayed response, and inconsistent recovery methods. This has led businesses to allocate resources to security awareness training initiatives and implement a wide range of security protocols, procedures, and technological measures [9].

III. AI-BASED INSIDER THREAT DETECTION APPROACHES

Insider threat detection has become an important application area for AI. Various AI techniques have been explored to analyze user activities, communications, and patterns to identify potential insider threats. The following subsections explore aspects of AI-driven insider threat detection strategies, each providing targeted insights into the complex field of ensuring organizational security. The sophisticated methodologies and technological advancements discussed consist of user behavior analytics and anomaly detection, Natural Language Processing (NLP) techniques, the transformative influence of Large Language Models (LLMs), and the graph-based approaches employed to discern insider threats. These innovative AI methods analyze user behavior, textual data, language patterns, and diverse datasets, contributing to evolving cybersecurity defenses against insider threats.

A. User Behavior Analytics and Anomaly Detection

Anomalies are data patterns that deviate from normal behavior. They can be caused by various factors, such as fraud, hacking, terrorism, or system failures. Although the causes differ, anomalies share the trait of being noteworthy to analysts. Their relevance and interest in real-world situations make anomaly detection important [10-11]. Machine Learning (ML) can automatically create required models for anomaly detection based on the training data provided. This approach is motivated by the availability of necessary training data, which are easier to obtain than manually defining models. As attacks become more complex, ML enables the building and maintenance of anomaly detection systems with minimal human input [12].

User Behavior Analysis (UBA) is a cybersecurity tool that detects insider threats, targeted attacks, and financial fraud. UBA examines human behavior patterns and uses ML algorithms and statistical analysis to identify potential threats based on abnormal deviations from normal activities [13]. User behavior profiles utilize a user's historical activities within an organization to reflect their typical actions and psychological characteristics. Malicious insiders may disguise themselves as

legitimate users, quickly logging into security domains to conduct malicious activities, such as stealing files, before logging out again. Using user profiles helps detect anomalous behaviors that deviate from normal activities, signaling potential insider threats [14].

B. Natural Language Processing (NLP)-based Approaches

NLP techniques are well suited to analyze textual data, such as emails, chats, social media posts, and documents, to identify potential insider threats. NLP can extract key information from unstructured text data to detect signs of disgruntlement, malicious intent, or unauthorized sharing of sensitive information. In [15], a novel method was introduced to detect insider threats through anomaly detection, using NLP techniques to discern unusual patterns within textual data. The findings emphasized the effectiveness of this approach in uncovering insider threats that conventional methods struggle to identify.

C. Large Language Models (LLMs)

Recent advances in NLP have led to the development of LLMs with hundreds of billions of parameters or more, trained on massive textual datasets. These LLMs, based on transformer architecture, demonstrate impressive capabilities in language understanding and the completion of complex tasks through text generation [16]. Recent advances in LLMs, such as Bidirectional Encoder Representations from Transformers (BERT), Generative Pre-trained Transformer (GPT) 4, and Google's Language Model for Dialogue Applications (LaMDA), present new opportunities for insider threat detection through NLP. LLMs can identify unusual patterns in the language used by insiders, such as the use of specific keywords or phrases associated with malicious intent. In [17], a pre-trained LLM was specifically developed to detect cybersecurity threats. This LLM could recognize 14 distinct attack categories, achieving an impressive accuracy rate of 98% across all identified attacks.

D. Graph-based Approaches

As technology has progressed, user data have become diverse and multidimensional, stemming from a range of sources including network activity, psychological elements, organizational dynamics, and employee conduct. These data exhibit distinct structural patterns. A graph-based method can be used to identify insider threats within this complex and heterogeneous dataset [18]. In [19], the challenge of insider threats within organizational contexts was addressed by proposing a method to detect insider behavior among employees by combining self-anomaly detection, which examines an employee's historical activities, with group anomaly detection, which compares an employee's behavior to that of their peers. This approach involved creating a contextual graph to provide relevant information for analysis. Experimental results demonstrated the algorithm's effectiveness in identifying insider instances and reducing false alarms when compared to using self-anomaly detection alone.

Table I presents a comparative analysis of insider threat detection, demonstrating how it benefits significantly from diverse AI-based approaches. Therefore, organizations might achieve significant improvements in the early detection and prevention of insider threats.

TABLE I. COMPARATIVE ANALYSIS OF APPROACHES FOR INSIDER THREAT DETECTION

Reference	Threat detection approach	Technique	Dataset	Accuracy metrics	Limitation
[20]	User behavior analytics for anomaly detection	LSTM autoencoder	CERT v4.2	True Positive Rate, False Positive Rate and Accuracy	Potential for missing important details when extracting features
[21]	User behavior analytics for anomaly detection	OCSVM, RNN, and Isolation Forest	Collected data from 4 employees of the organization	Precision, Recall, and Accuracy	Dataset with limited features
[22]	User behavior analytics for anomaly detection	Auto-encoder neural network, K-means clustering, and hidden Markov model	Credit card transactions by European cardholders	True Positive Rate, False Positive Rate	Limited dataset
[23]	NLP	Logistic regression and decision tree	Cyberbullying tweets	Accuracy	Need to investigate more methods for comparison
[15]	NLP	K-means and PCA	CERT r6.2	True Positive Rate	Lack of evaluation with real-world dataset
[17]	LLMs	SecurityLLM	Collected data from IoT devices	Accuracy, Recall, and F1-Score	Lack of different types of attacks
[24]	LLMs	Statistical analysis	Collected data from DC inside	Accuracy	Limited dataset
[25]	LLMs	Human-in-the-loop machine learning	Collected data from the communication of electrical devices	True Positive Rate, False Positive Rate, Precision and F1-Score	Low accuracy rate and lack of fine-tuning
[26]	Graph-Based Approach	Multi-edge weight relational graph neural network	CERT r6.2	Accuracy, Recall, Precision, and F1-Score	High delay in threat detection
[27]	Graph-Based Approach	Isolation Forest Algorithm	Collected data from individual users	Percentage of suspicious user	Limited features and dataset
[28]	Graph-Based Approach	Graph theoretic	Enron email dataset	Accuracy	Difficult to detect anomalies in dynamic data

IV. CHALLENGES AND LIMITATIONS OF DETECTING INSIDER THREATS

Detecting malicious insider threats poses significant challenges, as these events are often rare and difficult to model using only past data. Additionally, privacy concerns regarding employee monitoring and insufficient labeled data are critical issues for AI success. To train an AI model, large amounts of data are required. Thus, the model recognizes patterns, makes accurate predictions, and improves its performance. In this context, AI is used to model human behavior to identify insider threats. The challenges and limitations of detecting insider threats are as follows.

- Insufficient Insider Threat Datasets: Current research on insider threats struggles to validate and improve detection algorithms due to insufficient real-world organizational data. The lack of true insider threat data is a significant impediment to the development and evaluation of detection systems. Existing studies highlight that artificially generated datasets are not specifically designed to target insider threats. Furthermore, related research studies frequently use artificially generated datasets that are not suitable for insider threat scenarios. For example, some datasets lack malicious data or are out of date [3, 29].
- Privacy Concerns Related to Employee Monitoring: The solution proposed against insider threats should balance organizational security needs with employees' privacy rights and ethical concerns. For this purpose, more research is needed to develop monitoring strategies that are effective and respectful of employee autonomy and legal protections. The main problem with this concept arises from concerns about the employer's rationale for surveillance and its effects on morale and public perception [30].
- Budget and Time Issues: Supervised ML requires training data to build classification models for detecting insider threats. Existing approaches are mostly based on supervised learning that requires contextual user data and training procedures specific to insider threat detection. Although supervised ML methods are effective, they can be expensive and time-consuming [29].
- •Static Access Control Policy Rules: Traditional access control systems have inherent vulnerabilities due to their reliance on predefined, inflexible policies and static attributes for making authorization decisions. These legacy systems fail to detect critical dynamic security events, such as changes in user behavior patterns, abuse of granted privileges, compromise of credentials through theft of access cards or passwords, and structural modifications of protected data [31].
- Changing Employee Behavior: Determining the threshold of deviant behavior that constitutes a malicious insider, or distinguishing normal from abnormal actions when a bad actor is already present, poses challenges. Employee routines fluctuate with shifting work demands and coverages. Therefore, abnormal behaviors can be mislabeled as malicious. Thus, careful management is necessary to avoid false accusations against employees, reflecting the difficulties in building robust insider threat defenses [32].
- Collusion Attacks: In collusion attacks, two or more malicious insiders coordinate to compromise a system. Therefore, it is difficult to detect insider threats. The actions of each individual may seem benign on their own. However, insiders can gradually manipulate and steal data without raising red flags by avoiding large and abnormal actions of any user. To detect these attacks, monitoring must go beyond the tracking of individuals to identify collective suspicious patterns of activity on sensitive resources [33].

V. FUTURE DIRECTIONS AND RESEARCH OPPORTUNITIES

AI-based insider threat detection systems have made significant advances recently, and various fields warrant further research and exploration. Several key research avenues emerge to address the future directions of AI-based insider threat detection. First, building strong defenses is imperative due to the threat posed by adversarial attacks. To strengthen the resilience of AI systems against insider threats, adversarial attacks - which aim to circumvent or manipulate AI models - require concentrated efforts on strategies such as adversarial training, robust feature extraction, and anomaly detection in adversarial settings [34]. Improving the explainability of AI models used for insider threat detection is another important path. Researchers are urged to investigate methods that provide comprehensible justifications for the choices made by AI systems. Explainable AI models are expected to promote trust development, increase transparency, and enable productive cooperation between AI systems and human analysts [35]. Additionally, one promising direction is the integration of multimodal data analysis. To improve insider threat detection's precision and efficacy, multiple data modalities should be combined, including text, audio, and video. Future studies should focus on creating AI models that are adept at interpreting and combining data from various sources in a fluid and efficient manner. Insider threat detection through a thorough analysis of multimodal data is possible with techniques such as multimodal fusion, cross-modal learning, and deep learning architectures [36]. The development of privacy-preserving techniques becomes critical as the importance of privacy in insider threat identification increases. To effectively detect insider threats and protect employee privacy, research should focus on the analysis of cryptographic protocols, secure multiparty computation, and differential privacy. Ensuring that these techniques not only maximize efficiency and benefits but also unquestionably protect sensitive data integrity throughout the entire detection process is crucial [37-38]. Finally, human-centric methods that enable cooperation between AI systems and human analysts represent an important field for future study. It is crucial to investigate ways to incorporate contextual knowledge and human expertise into AI systems. Effective collaboration between human analysts and AI can be facilitated by using human-in-the-loop approaches, interactive visualization techniques, and decision support systems [39].

In summary, addressing these directions and research opportunities will be essential for future work in AI-based insider threat detection. Promising AI techniques that have been presented include LLMs, NLP, graph-based approaches, user behavior analytics, and anomaly detection. Continuous research and development will lead to strong and efficient insider threat detection systems, enhancing organizational security against changing cyber threats, despite obstacles such as limited datasets and privacy concerns.

VI. CONCLUSION

Insider threats pose significant risks to organizations. In this context, detecting and mitigating these threats is crucial for ensuring the confidentiality, integrity, and availability of information systems. This study presents an overview of the current AI techniques utilized for insider threat detection and highlights research challenges and opportunities in the field. AI techniques, such as user behavioral analytics and anomaly detection, NLP, LLMs, and graph-based approaches, have shown promise in identifying patterns and anomalies in user behavior that may indicate malicious intent. These techniques leverage ML algorithms, statistical analysis, and textual analysis to detect insider threats and potential indicators of malicious activities. However, several challenges and limitations must also be addressed to improve the effectiveness of insider threat detection. The main difficulty is the lack of sufficient datasets to detect insider threats. This deficiency hinders the development and evaluation of detection systems. In addition, privacy concerns related to employee monitoring and the need to balance organizational security needs with employee privacy rights and ethical considerations require further research. In this context, the development of monitoring strategies that respect privacy and legal protections is an essential requirement. Furthermore, the process of training accurate models to detect insider threats could be expensive and time-consuming, especially in supervised learning approaches that require contextual user data. Static access control policy rules and the changing behavior of employees pose additional challenges in accurately identifying insider threats and distinguishing them from normal actions.

Future directions of advanced AI-based insider threat research include the exploration of ensemble models, context-aware techniques, adversarial machine learning, and federated learning approaches. These approaches enable improving the accuracy and efficiency of insider threat detection systems.

In conclusion, insider threats present ongoing challenges for organizations, and the application of AI techniques holds promise in improving the early detection and prevention of such threats. Therefore, continued research and development in this field will contribute to the advancement of robust and effective insider threat detection systems, and ultimately improve the security posture of organizations against the evolving cybersecurity threats.

REFERENCES

[1] J. R. C. Nurse *et al.*, "Understanding Insider Threat: A Framework for Characterising Attacks," in *2014 IEEE Security and Privacy Workshops*, San Jose, CA, USA, May 2014, pp. 214–228, https://doi.org/10.1109/SPW.2014.38.

[2] "Cyber security breaches survey 2023," Department for Science, Innovation & Technology, London, UK. [Online]. Available: https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2023/cyber-security-breaches-survey-2023.

[3] I. Homoliak, F. Toffalini, J. Guarnizo, Y. Elovici, and M. Ochoa, "Insight Into Insiders and IT: A Survey of Insider Threat Taxonomies, Analysis, Modeling, and Countermeasures," *ACM Computing Surveys*, vol. 52, no. 2, Dec. 2019, https://doi.org/10.1145/3303771.

[4] T. E. Senator *et al.*, "Detecting insider threats in a real corporate database of computer usage activity," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, Chicago, IL, USA, May 2013, pp. 1393–1401, https://doi.org/10.1145/2487575.2488213.

[5] "Defining Insider Threats," *CISA*, https://www.cisa.gov/topics/physical-security/insider-threat-mitigation/defining-insider-threats.

[6] "2023 Data Breach Investigations Report," *Verizon*, https://www.verizon.com/business/resources/reports/dbir/.

[7] "2022 Cost of Insider Threats Global Report," Ponemon Institute, North Traverse City, MI, USA, 2022. [Online]. Available: https://www.proofpoint.com/sites/default/files/threat-reports/pfpt-us-tr-the-cost-of-insider-threats-ponemon-report.pdf.

[8] M. Theis *et al.*, "Common Sense Guide to Mitigating Insider Threats, Sixth Edition," Carnegie Mellon University, report, Sep. 2020, https://doi.org/10.1184/R1/12363665.v1.

[9] A. Georgiadou, S. Mouzakitis, and D. Askounis, "Detecting Insider Threat via a Cyber-Security Culture Framework," *Journal of Computer Information Systems*, vol. 62, no. 4, pp. 706–716, Jul. 2022, https://doi.org/10.1080/08874417.2021.1903367.

[10] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, Apr. 2009, https://doi.org/10.1145/1541880.1541882.

[11] M. H. H. Khairi, S. H. S. Ariffin, N. M. A. Latiff, A. S. Abdullah, and M. K. Hassan, "A Review of Anomaly Detection Techniques and Distributed Denial of Service (DDoS) on Software Defined Network (SDN)," *Engineering, Technology & Applied Science Research*, vol. 8, no. 2, pp. 2724–2730, Apr. 2018, https://doi.org/10.48084/etasr.1840.

[12] S. Omar, A. Ngadi, and H. H. Jebur, "Machine Learning Techniques for Anomaly Detection: An Overview," *International Journal of Computer Applications*, vol. 79, no. 2, pp. 33–41, Oct. 2013, https://doi.org/10.5120/13715-1478.

[13] T. Akutota and S. Choudhury, "Big Data Security Challenges: An Overview and Application of User Behavior Analytics," *International Research Journal of Engineering and Technology*, vol. 4, no. 10, pp. 1544–1548, Oct. 2017.

[14] X. Wang, Q. Tan, J. Shi, S. Su, and M. Wang, "Insider Threat Detection Using Characterizing User Behavior," in *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, Guangzhou, China, Jun. 2018, pp. 476–482, https://doi.org/10.1109/DSC.2018.00077.

[15] N. Garba, S. Rakshit, C. D. Mang, and N. R. Vajjhala, "An email content-based insider threat detection model using anomaly detection algorithms," in *Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021*, Apr. 2021, https://doi.org/10.2139/ssrn.3833744.

[16] W. X. Zhao *et al.*, "A Survey of Large Language Models." arXiv, Nov. 24, 2023, https://doi.org/10.48550/arXiv.2303.18223.

[17] M. A. Ferrag *et al.*, "Revolutionizing Cyber Threat Detection with Large Language Models: A privacy-preserving BERT-based Lightweight Model for IoT/IIoT Devices." arXiv, Feb. 08, 2024, https://doi.org/10.48550/arXiv.2306.14263.

[18] R. Nasir, M. Afzal, R. Latif, and W. Iqbal, "Behavioral Based Insider Threat Detection Using Deep Learning," *IEEE Access*, vol. 9, pp. 143266–143274, 2021, https://doi.org/10.1109/ACCESS.2021.3118297.

[19] Pratibha, J. Wang, S. Aggarwal, F. Ji, and W. P. Tay, "Learning Correlation Graph and Anomalous Employee Behavior for Insider Threat Detection," in *2018 21st International Conference on Information Fusion (FUSION)*, Cambridge, UK, Jul. 2018, pp. 1–7, https://doi.org/10.23919/ICIF.2018.8455358.

[20] B. Sharma, P. Pokharel, and B. Joshi, "User Behavior Analytics for Anomaly Detection Using LSTM Autoencoder - Insider Threat Detection," in *Proceedings of the 11th International Conference on Advances in Information Technology*, Bangkok, Thailand, Jul. 2020, pp. 1–9, https://doi.org/10.1145/3406601.3406610.

[21] X. Xi *et al.*, "An Ensemble Approach for Detecting Anomalous User Behaviors," *International Journal of Software Engineering and Knowledge Engineering*, vol. 28, no. 11–12, pp. 1637–1656, Nov. 2018, https://doi.org/10.1142/S0218194018400211.

[22] I. I. M. Abu Sulayman and A. Ouda, "User Modeling via Anomaly Detection Techniques for User Authentication," in *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Vancouver, BC, Canada, Oct. 2019, pp. 0169–0176, https://doi.org/10.1109/IEMCON.2019.8936183.

[23] T. Kanan, S. Hendawi, S. AlZu'bi, M. Elbes, and A. Mughaid, "Revolutionizing Cyberbullying Prevention: A Cutting-Edge Natural Language Processing-Based Approach," in *2023 International Conference on Information Technology (ICIT)*, Amman, Jordan, Aug. 2023, pp. 220–225, https://doi.org/10.1109/ICIT58056.2023.10225847.

[24] T. Kwon and C. Kim, "Efficacy of Utilizing Large Language Models to Detect Public Threat Posted Online." arXiv, Dec. 29, 2023, https://doi.org/10.48550/arXiv.2401.02974.

[25] A. Zaboli, S. L. Choi, T. J. Song, and J. Hong, "ChatGPT and other Large Language Models for Cybersecurity of Smart Grid Applications." arXiv, Nov. 09, 2023, https://doi.org/10.48550/arXiv.2311.05462.

[26] J. Xiao, L. Yang, F. Zhong, X. Wang, H. Chen, and D. Li, "Robust Anomaly-Based Insider Threat Detection Using Graph Neural Network," *IEEE Transactions on Network and Service Management*, vol. 20, no. 3, pp. 3717–3733, Nov. 2022, https://doi.org/10.1109/TNSM.2022.3222635.

[27] A. Gamachchi, L. Sun, and S. Boztas, "A Graph Based Framework for Malicious Insider Threat Detection." arXiv, Sep. 01, 2018, https://doi.org/10.48550/arXiv.1809.00141.

[28] W. Eberle, J. Graves, and L. Holder, "Insider Threat Detection Using a Graph-Based Approach," *Journal of Applied Security Research*, vol. 6, no. 1, pp. 32–81, Dec. 2010, https://doi.org/10.1080/19361610.2011.529413.

[29] M. N. Al-Mhiqani *et al.*, "A Review of Insider Threat Detection: Classification, Machine Learning Techniques, Datasets, Open Challenges, and Recommendations," *Applied Sciences*, vol. 10, no. 15, Jan. 2020, Art. no. 5208, https://doi.org/10.3390/app10155208.

[30] F. L. Greitzer, "Insider Threats: It's the HUMAN, Stupid!," in *Proceedings of the Northwest Cybersecurity Symposium*, Richland, WA, USA, Dec. 2019, https://doi.org/10.1145/3332448.3332458.

[31] M. Raissi-Dehkordi and D. Carr, "A multi-perspective approach to insider threat detection," in *2011 - MILCOM 2011 Military Communications Conference*, Baltimore, MD, USA, Nov. 2011, pp. 1164–1169, https://doi.org/10.1109/MILCOM.2011.6127457.

[32] N. Saxena, E. Hayes, E. Bertino, P. Ojo, K. K. R. Choo, and P. Burnap, "Impact and Key Challenges of Insider Threats on Organizations and Critical Businesses," *Electronics*, vol. 9, no. 9, Sep. 2020, Art. no. 1460, https://doi.org/10.3390/electronics9091460.

[33] I. A. Gheyas and A. E. Abdallah, "Detection and prediction of insider threats to cyber security: a systematic literature review and meta-analysis," *Big Data Analytics*, vol. 1, no. 1, Aug. 2016, Art. no. 6, https://doi.org/10.1186/s41044-016-0006-0.

[34] N. Papernot and P. McDaniel, "Deep k-Nearest Neighbors: Towards Confident, Interpretable and Robust Deep Learning." arXiv, Mar. 13, 2018, https://doi.org/10.48550/arXiv.1803.04765.

[35] S. S. Taher, S. Y. Ameen, and J. A. Ahmed, "Advanced Fraud Detection in Blockchain Transactions: An Ensemble Learning and Explainable AI Approach," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12822–12830, Feb. 2024, https://doi.org/10.48084/etasr.6641.

[36] T. Baltrušaitis, C. Ahuja, and L. P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, Jan. 2018, https://doi.org/10.1109/TPAMI.2018.2798607.

[37] M. Abadi *et al.*, "Deep Learning with Differential Privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, Vienna, Austria, Oct. 2016, pp. 308–318, https://doi.org/10.1145/2976749.2978318.

[38] P. Geetha, C. Naikodi, and L. Suresh, "Optimized Deep Learning for Enhanced Trade-off in Differentially Private Learning," *Engineering, Technology & Applied Science Research*, vol. 11, no. 1, pp. 6745–6751, Feb. 2021, https://doi.org/10.48084/etasr.4017.

[39] M. R. Endsley, "From Here to Autonomy: Lessons Learned From Human–Automation Research," *Human Factors*, vol. 59, no. 1, pp. 5–27, Feb. 2017, https://doi.org/10.1177/0018720816681350.