Journal of Interdisciplinary Science &Technology

Enhancing Cybersecurity with AI: Predictive Threat Detection Using LLMs

Prem Namdeo Kanade*, Satyam Bansidhar Adhav, Om Praful Dhoble
Computer Engineering, Sandip Institute of Technology and Research Center
Computer Engineering, Sandip Institute of Technology and Research Center
Computer Engineering, Sandip Institute of Technology and Research Center
premkanade27@gmail.com, *satyamaadhav@gmail.com, omdhoble2005@gmail.com*

Abstract: The rising frequency and complexity of cyber threats, including phishing, malware, and social engineering attacks, have made traditional cybersecurity methods are not enough. This paper explores how Artificial Intelligence (AI), especially predictive models and Large Language Models (LLMs), can improve threat detection and response. Predictive AI techniques, such as supervised and unsupervised learning, help identify anomalies and potential breaches by learning from historical and live network data. At the same time, LLMs provide strong natural language understanding. This ability allows for detecting phishing attempts, analyzing complex malware scripts, and extracting threat intelligence from unstructured text sources. This study gives a broad overview of AI-based cybersecurity methods, focusing on real-world applications, effectiveness, and integration challenges. It also addresses limitations like Weak spots that attackers can trick, explainability, and ethical issues. The findings show the transformative potential of AI in creating flexible, smart, and scalable cyber defense systems.

Keywords: Machine Learning, Large Language Models, Artificial Intelligence, Cybersecurity

----------∴ ∴ ∴----------

I INTRODUCTION

With the rise of new technologies and the expansion of the Internet of Things (IoT), the scope of CFA has diversified. Additional types of threats are hacking wirelessly or infiltrating smart homes and smart buildings. These new technologies allow users to perform daily tasks from anywhere, resulting in convenience. However, Global IT Security firm Cyber security Ventures has estimated the damage of cybercrimes to greatly exceed $6 trillion in 2021 and has growth prospects to further significantly exceed $20 trillion by the year 2023[1]. Advanced persistent threats or cyber espionage infiltrate an individual or organization configuration and capture or damage data.

With the current configuration of an organization, the firm tends to further build a detection system. Most security control systems are now based on the theory of systems. Security enforcement devices based on alarm systems create barriers. These barriers protect or restrict access to the premises; the intrusion detection devices are used. Security devices perform detection and alarm signaling as a complex automated system. With the current configuration of an organization, threats to an organization have expanded greatly. Most security control monitoring systems are now able to determine and maintain operating rules which are separate from the system controlling breaks to the system security devices [1], [2].

One of the new frontiers of AI is represented by Large Language Models such as ChatGPT, GPT-4, and other transformer-based models, as they seem to perform remarkably well in understanding and processing human language[8]. They can be trained on large sets of unstructured data such as emails, intelligence reports, or even entire code repositories. In addition, they can respond to queries in a contextual manner, identifying intricate language cues to assist in phishing, malware script analysis, and even social engineering mitigation.

This paper seeks to study how AI technologies, specifically LLMs and predictive models, can be integrated within a cybersecurity framework. It examines the methodologies, datasets, and technologies that were employed to build intuitive threat detection systems. It also analyzes the risks and benefits of LLMs. As a matter of emphasis, the paper seeks to reinforce the relationship between artificial intelligence and cybersecurity by demonstrating how the technologies can be employed to build powerful, autonomous, and adaptable systems to counter emerging cyber challenges.

*Corresponding Author- **Prem Namdeo Kanade**

II LITERATURE REVIEW

The incorporation of Artificial Intelligence (AI) into cybersecurity has garnered considerable attention both in academia and industry over the past ten years. Various research papers have been published on the use of ML, DL, and more recently, LLM for mitigation and detection of cyber threats. This section summarizes the most important works in the field while depicting the progress of AI-powered cybersecurity approaches.

In earlier studies, the use of more classical ML methods, like decision trees, SVMs, and random forests for a portion of IDS, garnered significant attention[2], [6]. These methods were trained to differentiate between malicious and benign network traffic using structured datasets like KDD99 and NSL-KDD. The methods were useful to a degree, but the static nature of the models made it difficult to handle new and novel attack dimensions.

The field of deep learning has significantly accelerated progress in the field of anomaly detection. RNNs and LSTMs have successfully identified sophisticated attack patterns in networks for years[3], [4]. Through the use of CNNs, the classification of malware has been made easier, as binary files can now be transformed into in detecting insider threats and zero-day attacks[5], the focus has lately moved to unsupervised and semi-supervised learning techniques. Approaches such as autoencoders and clustering algorithms Isolation Forest and DBSCAN have proven useful to some extent for outlier and behavioral anomaly detection without the use of labeled information [6].

Newer transformer-based LLMs have also created additional opportunities for work within the scope of cybersecurity. As seen in [7] and [8], phishing detection has been attempted using GPT-based models where they could detect the malicious intent hidden within verbosity of some emails and text messages. Other authors have demonstrated that LLMs can assist in malware script reverse engineering, as well as in summarizing threat intelligence from unstructured data like forum posts and blogs on the dark web and on security [9].

In addition, new frameworks have integrated AI technologies into real time SIEM as well as into SOAR systems. AI in these frameworks focuses on enhancing alert triaging, incidence response workflows, automation of low-level tasks and reducing false positive rates [10].

Regardless of the progress made in the domain, multiple issues are still present. Other researchers have shown that AI models are vulnerable to adversarial approaches, data poisoning, and biases [11], [12]. Furthermore, the absence of explanation in deep models

III METHODOLOGY

This research takes a hybrid approach utilizing traditional predictive machine learning models and transformer-based large language models (LLMs) to discover and respond to digital threats including, but not limited to phishing, malware, and social engineering attacks. The method consists of four stages: data collection, model selection and training, predictive AI for threat detection, and LLMs for contextual understanding.

A. Data Collection and Preprocessing

Our researcher will use a combination of publicly available datasets and hypothetically generated datasets, the use of which will replicate various threat scenarios:

NSL-KDD and CICIDS 2017/2018 will be used for intrusion detection [2].

Phish Tank, Nazario Email Dataset, and Spam Assassin will be used for phishing and spam detection.

The EMBER dataset will be used for static malware classification using binary features [5].

All datasets will be preprocessed through normalization, tokenization (for text), and feature extraction. For inputs to the LLMs phishing emails and malware command-line scripts will be reformatted as natural language prompts.

B. Predictive AI Models for Threat Detection

Supervised and unsupervised learning algorithms have been trained to spot anomalies in structured data:

Random Forest and SVM models to classify network-based threats.

Autoencoders were developed to do unsupervised anomaly detection.

CNN and LSTM models to model sequential (or time-series) data from logs and packet flows.

Model performance over data sets can be evaluated against standard metrics, including accuracy, precision, recall, F1-score, and ROC-AUC.

C. LLM-Based Threat Analysis

We leverage a fine-tuned version of GPT-4 to analyze textual data from suspected phishing emails, social engineering messages, and potentially malicious malware scripts[8]. The model accomplishes the following:

- Phishing Detection: identification of aspects of persuasive language, spoofed sender information, and malicious intent.

- Script Interpretation: summarization of suspicious behavior derived from command line or PowerShell scripts.

- Threat Intelligence Extraction: summarization of unstructured cyber threat reports, while identifying Indicators of Compromise (IOCs) and Threats, Techniques, and Procedures (TTPs) within the document.

We use contextual embeddings and cybersecurity prompts with domain-specific tokens to increase relevancy.

D. Hybrid Model Integration

A rule-based engine integrates output from both predictive models and (LLMs). It has a layered approach:

Phase 1: ML models perform real-time classification of events on the network and endpoints.

Phase 2: Suspicious text records are sent to the LLM for semantic analysis.

Phase 3: The decision is made based on ensemble scoring and policy-driven thresholds for remediation actions.

This hybrid approach is intended to balance swiftness (ML models) with semantic depth (LLM), while minimizing false positive anomalies.

E. Experimental Setup

Platform: Python-based implementation leveraging Scikit-learn, TensorFlow, and HuggingFace Transformers.

Hardware: NVIDIA RTX 3090 GPU and 64 GB of RAM.

Validation: 10-fold cross-validation for the ML models and a sequence of prompt evaluation metrics and human verification for the LLM outputs.

IV CONCLUSION

As cyber security threats continue to grow in terms of sophistication, scale, and frequency, defense mechanisms based on rules are insufficient to provide adequate protection. This research examined the role of Artificial Intelligence, through predictive machine learning models and Large Language Models (LLMs), in use within cybersecurity systems for the purpose of identifying and reducing the impact of phishing, malware, and social engineering attacks .

The research demonstrated that supervised and unsupervised learning algorithms can detect known and unknown threats using the mechanisms of pattern recognition and anomaly detection. Additionally, the LLMs such as GPT-4 use a semantic layer of meaning that allows understanding of the human language, script obfuscation, and extracting the appropriate threat intelligence from unstructured sources. The hybrid framework that consisted of both algorithms significantly reduced false positives, identified threats more accurately and improved the model's ability to respond and understand the complex nature of incidental attacks, which are steeped in meaning and context.

Adopting AI within cybersecurity systems is not without its own challenges. The computational costs, concealment and opacity of modelling, the vulnerability of adversarial inputs, and ethical considerations regarding surveillance and infrastructure are important to recognize. Future research should approach improving analyzer explainability, efficiency and robustness in threat identification, and related areas of federated learning and privacy preserving techniques supported through varying aspects of privacy intervention and data security.

In summary, the combination of AI and cybersecurity can pave the way for building intelligent, adaptive and scalable threat detection systems. As LLMs and predictive modeling continue to improve, AI will be an essential tool for combatting evolving cyber threats.

V REFERENCE

[1] D. E. Denning, "An Intrusion-Detection Model," *IEEE Transactions on Software Engineering*, vol. SE-13, no. 2, pp. 222–232, Feb. 1987.

[2] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," *Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, Ottawa, ON, Canada, 2009, pp. 1–6.

[3] R. Shone, V. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 41–50, Feb. 2018.

[4] Y. Kim, "LSTM-based system for anomaly detection in network traffic," *Computer Communications*, vol. 163, pp. 120–127, Nov. 2020.

[5] N. Saxe and K. Berlin, "Deep neural network based malware detection using two-dimensional binary program features," *Proceedings of the 10th International Conference*

on Malicious and Unwanted Software (MALWARE), Fajardo, PR, 2015, pp. 11–20.

[6] H. Ringberg, A. Soule, J. Rexford, and C. Diot, "Sensitivity of PCA for traffic anomaly detection," *ACM SIGMETRICS Performance Evaluation Review*, vol. 35, no. 1, pp. 109–120, June 2007.

[7] S. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from URLs," *Expert Systems with Applications*, vol. 117, pp. 345–357, Mar. 2019.

[8] OpenAI, "GPT-4 Technical Report," *OpenAI*, 2023. [Online]. Available: https://openai.com/research/gpt-4

[9] T. Bhowmik, F. M. A. Rahman, and M. Hossain, "AI-based dark web threat intelligence collection and analysis," *Computers & Security*, vol. 97, p. 101951, Oct. 2020.

[10] IBM, "IBM QRadar SIEM with Watson: AI-powered threat detection and investigation," *IBM Security White Paper*, 2022. [Online]. Available: https://www.ibm.com/security

[11] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami, "Practical black-box attacks against deep learning systems using adversarial examples," *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, pp. 506–519, Apr. 2017.

[12] B. Biggio and F. Roli, "Wild patterns: Ten years after the rise of adversarial machine learning," *Pattern Recognition*, vol. 84, pp. 317–331, Dec. 2018

