

Cybersecurity Threat Hunting Through Log Analysis Using Large Language Models: A Comprehensive Review

Omar Salem Almutairi MidOcean University
Email: 251002148@st.midocean.eu.km

Abstract—This report offers an in-depth look at how Large Language Models (LLMs) are used in cybersecurity threat hunting, with a focus on log analysis. It discusses the significant potential of LLMs while also addressing the challenges they face and suggesting ways for future research and practical use. LLMs represent a major step forward in automating and improving log analysis for proactive threat detection, moving past the restrictions of traditional rule-based systems. Recent research shows promising methods, including better anomaly detection, efficient log parsing and embedding, sophisticated multi-agent collaboration, and new techniques for analyzing system states, which all perform better in specific areas.

However, there are still major challenges. These include traditional issues like scalability, the high number of false positives, and the difficulties of integrating LLMs into existing cybersecurity frameworks. Additionally, significant problems like vulnerabilities to adversarial attacks, such as prompt injection and data poisoning, data privacy risks from LLM operations, and important performance issues related to real-time responsiveness in changing cybersecurity environments need urgent attention. The lack of clarity in LLM decision-making also highlights the need for better interpretability and explainability.

To ensure the responsible and effective use of LLMs, this report stresses the importance of developing Human-in-the-Loop (HITL) systems, creating strong and standardized benchmarking frameworks, prioritizing Explainable AI (XAI) during the design phase, and conducting thorough cost-benefit analyses to guide investment and adoption choices. These steps are crucial for building trust, ensuring reliability, and enabling safe deployment of LLM-based solutions in real-world threat hunting efforts.

Index Terms—Large Language Models (LLMs), Cybersecurity, Threat Hunting, Log Analysis, Anomaly Detection, Explainable AI (XAI), Adversarial Attacks.

I. INTRODUCTION

THE current cyber threat landscape is marked by increasingly sophisticated attack methods, a growing volume of harmful activities, and highly dynamic threat actors. This situation requires a shift from reactive security measures, which respond to incidents after they happen, to proactive strategies like threat hunting. Threat hunting involves actively searching for potential threats that might have gotten past initial security defenses, aiming to spot malicious activities before they fully develop and cause serious damage. Traditional log analysis methods, which have long been a key part of cybersecurity, are becoming less effective in this changing environment. These methods typically rely on set rules and manual checks, making them slow, error-prone, and unable to detect new or changing threats. The large volume and complexity of log data produced

by modern IT systems further worsen these issues, making traditional approaches inefficient for thorough threat detection.

Large Language Models (LLMs) are well-suited to transform log analysis in cybersecurity due to their unique ability to handle and understand vast amounts of unstructured and semi-structured data. This skill allows LLMs to recognize subtle patterns, spot anomalies, and produce actionable insights that often exceed the capabilities of conventional methods. This paper reviews the current state of LLM applications in cybersecurity threat hunting, highlights critical challenges, and suggests future research and implementation strategies.

II. CORE APPLICATIONS AND METHODOLOGIES OF LLMs IN LOG ANALYSIS

LLMs' natural language processing (NLP) abilities are crucial for breaking down and understanding the complex and often chaotic nature of log data. This allows for the analysis of different log formats, content, and structures from various sources, including system, network, and application logs. Effectively processing and understanding this data is key to identifying subtle signs of compromise during threat hunting operations. It helps security analysts interpret seemingly unrelated log entries. Several methods and frameworks have been proposed to use LLMs for log analysis in threat hunting, mainly categorized into anomaly detection, log parsing, and multi-agent collaboration.

For example, **LogLLM** uses pre-trained LLMs like BERT and Llama to improve semantic understanding and anomaly detection. **LogRAG** is a semi-supervised framework that applies Retrieval-Augmented Generation (RAG) principles for staged detection, which significantly lowers false positives. In log parsing, **LLMeLog** enhances log events using in-context learning, achieving high F1-scores even with little labeled data. Similarly, a **hybrid approach** uses LLMs to generate REGEX and perform strong event embedding. New strategies like **Audit-LLM**, a multi-agent framework, improve log-based insider threat detection through the cooperation of agents and an Evidence-based Multi-agent Debate (EMAD) mechanism, which notably enhances explanation accuracy. Moving beyond traditional log analysis, methods like **"From Log to Fingerprint"** turn raw data into unique 'fingerprints' for better system state analysis, allowing for real-time anomaly detection and performance improvement. This cooperative integration shows that LLMs are strong tools that enhance

current deep learning methods, indicating a future of hybrid designs optimized for complex pattern recognition and natural language understanding.

A. The Fundamental Role of LLMs in Processing Diverse and Unstructured Log Data

Large Language Models are particularly suited for the challenges of cybersecurity log analysis due to their ability to process and comprehend natural language. Logs usually come in various semi-structured or completely unstructured formats, making traditional rule-based parsing and analysis inefficient and error-prone. LLMs, with their understanding of syntax, semantics, and context, effectively normalize, interpret, and extract meaningful insights from diverse log entries, regardless of their source (for instance, system events, network traffic, application logs). This skill allows security analysts to move beyond rigid keyword searches to a more flexible and intelligent system that can spot subtle anomalies and suspicious behaviors that might otherwise be missed. The large volume and speed of modern log data demand an automated and intelligent processing layer, which LLMs are particularly equipped to provide, supporting a comprehensive and flexible approach to threat hunting.

B. Detailed Examination of Key Frameworks

The research community has put forward several innovative frameworks that leverage LLMs for better log analysis in threat hunting. These frameworks often combine LLM strengths with established machine learning techniques to achieve better performance.

1) *Anomaly Detection*: Anomaly detection is central to proactive threat hunting. Its goal is to find deviations from typical system behavior that may signal malicious activity. Frameworks like **LogLLM** exemplify this by using pre-trained LLMs, particularly BERT and Llama, for enhanced semantic understanding and anomaly detection. This method has shown better performance than traditional deep learning approaches in capturing semantic information, leading to improved anomaly detection [1]. A notable example is **LogRAG**, a semi-supervised framework based on Retrieval-Augmented Generation principles. LogRAG conducts phased detection using both log tokens and templates, effectively reducing the impact of parsing errors. Its single-class classifier models normal system behavior, helping to cut down false positives [2].

2) *Log Parsing and Embedding*: Log parsing and embedding are key steps in preparing raw log data for analysis, where LLMs bring significant benefits. **LLMeLog** is a new approach to anomaly detection that uses LLMs to enhance log events through in-context learning techniques. It fine-tunes a pre-trained BERT model with these enriched log events and then trains a transformer-based anomaly detection model. Evaluations show that LLMeLog achieves high F1-scores even when using only a small percentage of labeled data for training, demonstrating its efficiency in data use [3]. Similarly, a **hybrid method** employs LLMs to create regular

expressions (REGEX) for accurate log parsing and uses a pre-trained LLM for strong event embedding [4]. This method also includes an autoencoder with an attention mechanism for unsupervised learning of log patterns, achieving high accuracy on public datasets for anomaly detection.

3) *Multi-Agent Collaboration*: Multi-agent collaboration is a growing and complex strategy in threat hunting. It involves coordinating multiple intelligent agents to tackle complicated tasks. **Audit-LLM** is a prime example designed for log-based insider threat detection [5]. It has three collaborating agents: the Decomposer, which breaks down complex tasks using Chain-of-Thought reasoning; the Tool Builder, which creates reusable tools for sub-tasks; and the Executor, which generates the final detection conclusion by utilizing created tools. The framework also includes a pair-wise Evidence-based Multi-agent Debate (EMAD) mechanism to refine conclusions through iterative reasoning exchange. Extensive experiments show its superiority over existing methods, with significant improvements in the trustworthiness of explanations from LLMs, directly addressing the need for clarity in LLM use.

4) *System State Analysis and Fingerprinting*: Going beyond traditional log analysis, this innovative strategy uses LLMs to create abstract, rich representations of system states for advanced change detection and anomaly identification. The **"From Log to Fingerprint"** approach turns raw log data and other system metrics into a unique "fingerprint"—a compact, informative representation that captures the essence of a complex system at any time [6]. This method enables intelligent compression, allowing for a new level of abstraction for system states. It uses multidimensional projection to map hundreds or thousands of metrics into an optimized multidimensional space, revealing hidden connections and dependencies. The generated fingerprints are designed for efficient storage and retrieval, which is essential for real-time analysis. The main innovation lies in comparing these fingerprints to instantly detect even the smallest differences between two system states, removing the need to manually review gigabytes of logs. This breakthrough allows for real-time anomaly detection, monitoring system drift, optimizing performance, and speeding up debugging and troubleshooting by providing a clear picture of changes that may have caused an issue. It represents a major shift in monitoring and managing complex systems, leading to greater stability, predictability, and effectiveness.

III. CRITICAL ASSESSMENT OF CHALLENGES AND LIMITATIONS

Even though LLMs have the potential to transform cybersecurity threat hunting, many significant challenges keep them from being widely and safely used. One major issue is **scalability**. LLMs have a hard time processing the growing amounts of log data quickly. This need for extensive computational resources can prevent many organizations from adopting them. Another problem is **false positives**. LLMs often misidentify harmless activities as threats. This can overwhelm security teams, causing 'alert fatigue' and wasting resources. Additionally, the complexity of **integration** is a big hurdle. It takes a lot of effort and possibly redesigning systems to smoothly

add LLM-based solutions to current cybersecurity setups like SIEM and SOAR platforms.

A major issue is the **interpretability and explainability (XAI)** of LLMs. Their 'black box' nature makes it hard for analysts to figure out why specific threats are flagged. This lack of understanding erodes trust and makes it difficult to validate and refine the models. There are also broader research findings that point to critical challenges. LLMs are very **vulnerable to adversarial attacks**. These include prompt injections that manipulate outputs, data poisoning that harms model integrity, and model inversion attacks that risk leaking sensitive data. Such attacks can allow threats to go undetected or lead to poor decision-making, significantly increasing false positives and negatives [7]. Additionally, concerns about **data privacy and ethical issues** are crucial. LLMs may unintentionally leak sensitive data from their training data, reinforce biases, and can 'hallucinate' inaccurate information, harming the credibility of threat intelligence [8]. Finally, **real-time performance and latency** are also vital issues. Even with advanced analytical skills, LLMs can be slow to respond, which is problematic in time-sensitive threat hunting situations where quick decisions are necessary [9]. Improving LLMs for low latency requires careful prompt design, effective token management, and possibly using distributed processing. This shows that these are not just technical difficulties, but fundamental challenges that need strong solutions for safe use.

A. In-depth Analysis of Core Challenges

Excitement about LLMs in cybersecurity should be balanced with a clear understanding of their key operational and inherent challenges. **Scalability** is still a tough barrier. The huge amount and speed of log data in modern IT environments often exceed the processing capability of LLMs, making real-time analysis labor-intensive and impractical for many organizations. The **false positive** issue worsens this situation. The probabilistic nature of LLMs can cause incorrect alerts that flood security teams with unnecessary notifications, leading to 'alert fatigue.' This, in turn, lowers the system's effectiveness and trust. Moreover, **integrating** new LLM solutions into existing, often outdated, cybersecurity systems like SIEM and SOAR is complicated. It requires significant development efforts for custom connectors and data pipeline redesigns. Perhaps most importantly, the lack of **interpretability and explainability (XAI)** in LLMs is a significant 'black box' problem [10]. Security analysts struggle to understand how LLMs reach specific conclusions, which damages trust, complicates debugging, and makes it hard to act on LLM-generated insights in critical situations where accountability matters.

B. Expanded Challenges from Broader Research

Beyond the obvious operational challenges, further research on LLMs uncovers deeper, more concerning issues that are especially relevant in cybersecurity. A major threat is their susceptibility to **adversarial attacks**. Techniques like prompt injection enable attackers to manipulate LLM inputs, bypass security measures, or create harmful outputs. Data poisoning can introduce biases or damage model integrity during

training. Model inversion attacks can lead to the leaking of sensitive data by inferring training data from LLM outputs. In threat hunting, these attacks can result in sophisticated evasion, allowing threats to go undetected, or lead to poor decision-making with dangerously high false positive rates [7]. Moreover, the nature of LLMs raises serious **data privacy and ethical issues**. Their training on large datasets carries risks of sensitive data leakage. These models can also unintentionally reinforce harmful biases present in their training data, resulting in unfair or incorrect threat evaluations. A well-known issue is **hallucinations**, where LLMs produce sounds plausible but entirely false information, which is disastrous in a field that demands complete accuracy [8]. Lastly, achieving **real-time performance and low latency** is a critical challenge that often gets overlooked. Although LLMs provide strong analytical capabilities, their high computational demands can slow response times. In fast-moving cybersecurity environments, where quick decision-making is key to preventing or handling threats, even slight delays can lead to missed chances or serious harm. This highlights that improving speed is just as essential as ensuring accuracy [9].

IV. FUTURE DIRECTIONS AND STRATEGIC RECOMMENDATIONS

The ongoing growth of large language models (LLMs) in threat hunting calls for clear future directions and recommendations to make the most of their potential and lessen associated risks. First, improving **interpretable and explainable LLM models (XAI)** is vital. Building trust among security analysts needs transparent models that clarify their operations and decision-making processes. This could involve using LLMs to convert complex outputs into clear narratives. Techniques like LIME are crucial for achieving local interpretability, directly addressing the 'black box' issue [10]. Second, creating **domain-specific and robust LLM architectures** is essential. While general-purpose LLMs show promise, fine-tuning them on relevant cybersecurity datasets will significantly improve their performance and make them more relevant for specific log analysis needs, aiming for deep domain-specific intelligence. Third, improving **scalability and real-time processing** is critical for managing large and complex log data. This requires exploring advanced techniques like distributed processing, parallel computing, optimized pre-training, and resolving latency challenges through concise prompt engineering and smart context management. This ensures quick decision-making in time-sensitive settings [9]. Fourth, integrating **Human-in-the-Loop (HITL)** systems is essential for better decision-making. HITL models weave human expertise into automated processes, allowing analysts to verify alerts, approve responses, and oversee actions. This connection is especially important in Security Operations Center (SOC) and Security Orchestration, Automation, and Response (SOAR) platforms [11]. This change is necessary, as it counters unreliable outputs and vulnerability to adversarial attacks. Fifth, **standardized benchmarking and evaluation** are crucial for objectively measuring progress, ensuring reproducibility, enabling fair comparisons, and guiding real-world implementation. Developing specific cybersecurity LLM

benchmarks that reflect the dynamic nature of threats, covering various log types and attack vectors, is needed for establishing reliability and trust [12]. Finally, addressing practical adoption requires thorough **cost-benefit analysis**. Organizations must weigh all associated costs against tangible and intangible benefits like risk reduction, compliance advantages, and savings from incident responses to justify strategic investments, considering uncertainties through sensitivity analysis [13].

V. CONCLUSION

Integrating large language models into cyber threat hunting, especially through log analysis, presents a powerful and transformative opportunity in cybersecurity. LLMs signal a shift away from traditional, often reactive, rule-based systems towards automated, proactive, and meaning-driven threat detection abilities. Current research showcases promising advancements across key applications, including sophisticated anomaly detection frameworks like LogLLM and LogRAG, efficient log parsing and embedding methods like LLMLog, innovative multi-agent systems like Audit-LLM, and new approaches to system state analysis like 'From Log to Fingerprint.' These methods have consistently shown better performance metrics across specific tasks, indicating that LLMs can greatly enhance threat intelligence and response capabilities in modern security operations.

Nevertheless, achieving reliable and secure deployment of LLM-based threat hunting solutions involves major, multifaceted challenges. These include not only the well-known issues of scalability and false positives, but also critical security risks like vulnerability to adversarial attacks, ethical dilemmas such as model bias and data privacy issues, and key operational constraints, like real-time response needs in fast-changing cybersecurity situations. To effectively navigate this complex landscape, the following recommendations are crucial for guiding future research and practical implementation efforts: a comprehensive approach that combines technical performance with strong security and ethical considerations; a human-centered design that focuses on interpretable and explainable LLM models and incorporates strong HITL systems to build trust and support human-AI collaboration; significant investment in creating and using standardized benchmarking frameworks that accurately mirror real-world complexities; conducting thorough cost-benefit analyses to provide clear cases for investment; and developing an adaptive security posture around these AI systems to counter evolving adversarial techniques. Following these guidelines will allow for the secure, efficient, and responsible use of LLM-based solutions, ultimately improving proactive cyber defense.

REFERENCES

- [1] S. Akhtar, S. Khan, and S. Parkinson, "LLM-based event log analysis techniques: A survey," *arXiv preprint arXiv:2502.00677*, 2025.
- [2] L. Zhang, T. Jia, M. Jia, and Y. Li, "LogRAG: Semi-Supervised Log-based Anomaly Detection with Retrieval-Augmented Generation," *ResearchGate*, 2025.
- [3] A. Fariha, V. Gharavian, M. Makrehchi, S. Rahnamayan, S. Alwidian, and A. Azim, "Log Anomaly Detection by Leveraging LLM-Based Parsing and Embedding with Attention Mechanism," *ResearchGate*, 2024.
- [4] S. Akhtar, S. Khan, and S. Parkinson, "LLM-based event log analysis techniques: A survey," *arXiv preprint arXiv:2502.00677*, 2025.
- [5] C. Song, L. Ma, J. Zheng, J. Liao, H. Kuang, and L. Yang, "Audit-LLM: Multi-Agent Collaboration for Log-based Insider Threat Detection," *arXiv preprint arXiv:2408.08902*, 2024.
- [6] M. R. Hawthorne, S. L. Plotkin, and B. Douglas, *Fingerprints: Analysis and Understanding the Science*, 2nd ed. Routledge, 2021.
- [7] E. Shayegani, M. A. A. Mamun, Y. Fu, P. Zaree, Y. Dong, and N. B. Abughazaleh, "Survey of Vulnerabilities in Large Language Models Revealed by Adversarial Attacks," *arXiv preprint arXiv:2310.10844*, 2023.
- [8] Analytics Vidhya, "Exploring Ethics and Privacy in the World of Advanced Language Models," *Analytics Vidhya Blog*, 2023.
- [9] M. Q. Li and B. C. M. Fung, "Security Concerns for Large Language Models: A Survey," *arXiv preprint arXiv:2501.04437*, 2025.
- [10] E. Kate, "Explainable AI (XAI) in Cybersecurity: Bridging the Gap Between AI and Human Understanding," *ResearchGate*, 2025.
- [11] International Journal of Multidisciplinary Scientific Research and Development, "Human-in-the-Loop Intelligence: Advancing AI-Centri
- [12] M. R. Hawthorne, S. L. Plotkin, and B. Douglas, *Fingerprints: Analysis and Understanding the Science*, 2nd ed. Routledge, 2021.
- [13] M. R. Hawthorne, S. L. Plotkin, and B. Douglas, *Fingerprints: Analysis and Understanding the Science*, 2nd ed. Routledge, 2021.