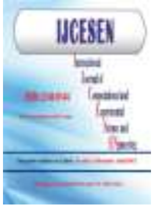
International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 10-No.3 (2024) pp. 507-516
http://www.ijcesen.com

Copyright © IJCESEN

ISSN: *2149-9144*



Review Article

A Comprehensive Review of Large Language Models in Cyber Security

Mesut GÜVEN*

TOBB University of Economics and Technology, TR-06560 Ankara, Turkey
* **Corresponding Author Email:** mesuttguven@gmail.com - **ORCID:** 0000-0002-0957-8541

Article Info:

DOI: 10.22399/ijcesen.469
Received : 24 September 2024
Accepted : 30 September 2024

Keywords :

Artificial Intelligence
Machine Learning
Large Language Models
Cyber Security
Malware Analysis

Abstract:

In response to the escalating complexity of cyber threats and the rapid expansion of digital environments, traditional detection models are proving increasingly inadequate. The advent of Large Language Models (LLMs) powered by Natural Language Processing (NLP) represents a transformative advancement in cyber security. This review explores the burgeoning landscape of LLM applications in cyber security, highlighting their significant potential across various threat detection domains. Recent advancements have demonstrated LLMs' efficacy in enhancing tasks such as cyber threat intelligence, phishing detection, anomaly detection through log analysis, and more. By synthesizing recent literature, this paper provides a comprehensive overview of how LLMs are reshaping cyber security frameworks. It also discusses current challenges and future directions, aiming to guide researchers and practitioners in leveraging LLMs effectively to fortify digital defences and mitigate evolving cyber threats.

1. Introduction

Given the escalating complexity of cyber threats and the expansive scope of digital environments, conventional security measures are increasingly challenged to keep pace with evolving adversarial tactics. Recent advancements in Artificial Intelligence (AI), particularly with large language models like GPT-3 and BERT, have opened new avenues for addressing these challenges. LLMs have demonstrated significant prowess in handling large-scale data analytics, performing multi-modal tasks, and generating content with remarkable fidelity [1]. Moreover, they exhibit capabilities in perception, cognition, and decision-making within AI tasks, making them formidable tools in cyber security [2].

The integration of LLMs into cyber security practices marks a transformative shift, particularly in threat detection and discovery within digital ecosystems. While systematic reviews have explored LLM applications in various fields of cyber security, there remains a notable gap in comprehensive discussions specifically focused on integrating LLMs into the cyber threat detection phase [3]. This review aims to address this gap by consolidating and analyzing representative applications of LLMs in cyber threat detection and discovery.

Aligned with the National Institute of Standards and Technology cybersecurity framework, which encompasses identification, protection, detection, response, and recovery phases, our review focuses on how LLMs can augment the detection phase. This involves leveraging their capabilities to enhance threat detection accuracy, mitigate false positives, and improve real-time response mechanisms.

The specific contributions of this work are respectively:

- This review provides a relatively exhaustive survey of LLMs utilized specifically to detect cyber threats, focusing on thematic works that can be applied in practical security scenarios.
- From the perspective of defenders, we focus on exploring how to use LLMs to enhance the detection of cyber threats.
- The research has categorized various stages of tasks in security scenarios that are closely correlated to LLMs. In this way, we

investigate the suitability of LLMs for specific security tasks and their potential roles across different stages.

- This review seeks to support the smooth incorporation of LLMs into existing security frameworks or the creation of domain-specific LLMs by security experts from diverse research fields.
- After comparing thematic works, we have compiled further analyses on LLMs and cyber threats from various scholarly sources and publications, synthesizing and presenting our perspectives on the potential advantages and obstacles of utilizing LLMs for detecting cyber threats and enhancing overall network security.

The structure of this review is as follows: Section 2 provides a brief historical overview connecting LLMs with the evolution of threat detection methodologies. Section 3 comprehensively reviews relevant literature on LLM applications in different security scenarios. Section 4 presents additional insights and discussions on related topics. Finally, Section 5 concludes the study.

2. Background

With the advancement of deep learning, LLMs were born in natural language processing. After that, some applications of LLMs have emerged in cybersecurity, notably in threat detection [4]. This section outlines some key milestones in the history of LLMs and cyberspace threat detection, shedding light on the inherent connections between these two domains.

With the scale of neural networks expanding, researchers have discovered that deep learning models exhibit advanced capabilities that traditional models cannot demonstrate. These abilities include sophisticated reasoning, sentiment analysis, and complex context abstraction [5]. This phenomenon, known as intellectual emergence, marks a crucial evolution in machine learning, distinguishing large models from their predecessors [6].

Among these large models, language models are particularly prominent in natural language processing applications, showcasing proficiency in various language-related tasks such as text generation, question answering, and intelligent information retrieval. To enhance their expressive power, LLMs like GPT-3 incorporate billions of parameters and employ efficient network architectures such as Transformers [7]. GPT-3, for instance, boasts over 175 billion parameters, enabling it to handle diverse linguistic contexts effectively. High-quality training data also plays a pivotal role in LLMs' effectiveness by broadening their knowledge base and improving generalization across tasks. Also, it is highlighted that GPT-3 was trained on a dataset comprising 45 terabytes of compressed text, filtered to approximately 570 GB, underscoring the importance of extensive data resources in model training [1].

Moreover, advancements in training methodologies and computational resources contribute significantly to the efficacy of LLMs. Institutions like Google and Meta have developed advanced LLMs such as PaLM2, Gemini, and Llama, which extend beyond traditional NLP tasks to applications in finance, healthcare, and customer service. The evolution of LLMs traces back to foundational NLP models like Word2Vec was one of the pioneers [8]. Word2Vec introduced distributed word representations, enhancing semantic understanding by embedding words into continuous vector spaces. This approach laid the groundwork for subsequent LLMs like BERT and GPT, which excel in capturing contextual nuances and semantic relationships. In parallel, advanced sequence modeling with the introduction of the gated recurrent unit, refining LSTM networks for improved language modeling capabilities [9]. These innovations have profound implications for LLMs, enabling them to model long-range dependencies and complex linguistic patterns more effectively. As time progressed, the availability of high-quality data and robust computing resources became more prevalent in AI research. Researchers increasingly focused on transfer learning and pre-training models. In 2018, two groundbreaking language models emerged: BERT developed by Google and GPT developed by OpenAI. BERT was introduced with a bidirectional transformer architecture, enabling the model to capture contextual information from both directions of word sequences. BERT demonstrated its capability to learn universal language patterns and assess contextual relevance through extensive pre-training on large datasets, followed by fine-tuning for specific tasks. Consequently, BERT has demonstrated exceptional performance across various tasks, including expert questions and answers, text classification, and named entity recognition [10].

Radford and Narasimhan implemented the auto-regressive Transformer architecture in GPT-1, which incorporates over 117 million parameters. Through a combination of unsupervised pre-training and supervised fine-tuning, GPT has shown proficiency in handling complex NLP tasks, particularly in text generation [11].

In reviewing the history of LLMs, it is evident that BERT and GPT have played milestone roles and represented two prominent pathways for subsequent LLM developments, as illustrated in Table 1.

Table 1. *Milestones in the Development of Large Language Models.*

Year	Model	Description
2018	BERT	Bi-directional transformer architecture for contextual language understanding
2018	GPT-1	Auto-regressive Transformer architecture for text generation

Building upon the foundational achievements of BERT and GPT, subsequent developments have cemented these models as benchmarks in the realm of large language models (LLMs). OpenAI's rapid succession of GPT models, starting with GPT-2 in 2019 and followed closely by the groundbreaking GPT-3 in 2020, marked significant milestones in AI capabilities. These models, with their expansive parameter sizes and advanced Transformer architectures, have not only pushed the boundaries of natural language processing but also set new standards for complexity and performance in AI systems [12].

Moreover, OpenAI's introduction of InstructGPT in March 2022 underscored a shift towards more adaptive LLMs, utilizing reinforcement learning with human feedback. This approach has enabled models like GPT to excel in multitasking, inference, prediction, and zero-shot learning, aligning outputs more closely with human preferences. Concurrently, advancements in BERT variants such as CAN-BERT have expanded their utility across diverse NLP applications, from text processing to intelligent translation, enhancing their adaptability and performance in real-world contexts [13].

Despite these advancements, concerns surrounding the security and ethical implications of LLMs have emerged alongside their widespread adoption. Issues such as prompt injection, data poisoning attacks, and vulnerabilities to backdoor intrusions remain significant challenges [14, 15]. Efforts to mitigate these risks through enhanced security frameworks and rigorous testing protocols are underway, aiming to safeguard user data and uphold ethical standards [16].

Looking forward, the proliferation of LLMs in various sectors has sparked what some term the “hundred models war” reflecting intense competition and innovation among developers [17]. This competitive landscape has not only driven advancements in content generation, multi-modal processing, and semantic analysis but also democratized access to AI capabilities through user-friendly platforms like ChatGPT, which simplify the creation of custom LLMs with minimal coding requirements.

As the field continues to evolve, the development of both open-source and proprietary LLMs has synergistically advanced the state-of-the-art. Innovations in encoder-decoder architectures and domain-specific adaptations, such as SecurityGPT, highlight ongoing efforts to tailor LLMs to specific industry needs. These advancements, coupled with improvements in model optimization techniques and scalability, are making LLMs increasingly accessible to micro-enterprises and medium-sized enterprises, despite the computational demands associated with models like GPT-3 [18].

Since the inception of large language models like BERT and GPT in the 2010s, their integration into cybersecurity has introduced transformative capabilities. These models represent a paradigm shift by leveraging their context-aware processing, extensive knowledge bases, and adaptive learning capabilities to address complex challenges in threat detection and discovery within cyberspace. Cyberspace threats encompass a broad spectrum of malicious activities, ranging from traditional malware to sophisticated cyber-attacks orchestrated by individual hackers, criminal organizations, and nation-states. These threats exploit technical vulnerabilities, system weaknesses, and human errors, necessitating advanced detection methodologies capable of mitigating risks effectively.

Early approaches in threat detection predominantly relied on signature-based and rule-based methods, which, although effective against known threats, struggled with adaptability to new and evolving attack vectors. The advent of machine learning brought significant advancements, dividing into supervised and unsupervised learning paradigms. Supervised learning techniques, such as support vector machines and naive bayes classifiers, excel in classifying threats based on labeled datasets, while unsupervised methods like K-means clustering provide flexibility in identifying anomalous patterns without prior labeling. However, both approaches face challenges in handling the non-linear and high-dimensional nature of cyber threats [19].

Deep learning has further revolutionized threat detection by harnessing neural networks like convolutional neural networks, Recurrent neural networks, and transformers. These models excel in capturing complex spatial and temporal relationships within network traffic data, enhancing the detection of advanced persistent threats [20]. Additionally, variants such as graph convolutional networks have demonstrated efficacy in analyzing network structures and relationships critical for identifying threat origins and propagation.

Reinforcement learning has emerged as a complementary technique, optimizing threat detection strategies through continuous learning and adaptation based on environmental feedback. This technique enhances autonomous threat detection by refining reward mechanisms and dynamically adjusting detection policies to mitigate emerging threats effectively. Furthermore, visualization tools and technologies rooted in computer vision facilitate intuitive analysis and identification of threat patterns within large-scale datasets, empowering security analysts with actionable insights [21].

The integration of cyberspace threat intelligence has augmented traditional threat detection by leveraging real-time and historical threat data to enhance situational awareness and predictive capabilities. Cyber threat intelligence enables automated threat identification and proactive defense measures against known and unknown threats, mitigating risks in dynamic cyber environments.

In conclusion, these advanced methodologies and technologies are propelling the evolution of threat detection towards automation, generalization, and integration with cloud computing and big data analytics [22]. While LLMs offer promising solutions to enhance detection accuracy and adaptability in complex cybersecurity landscapes, ongoing research is essential to address challenges such as model interpretability, scalability, and the evolving nature of cyber threats [23]. These advancements underscore the critical role of AI-driven approaches in fortifying cybersecurity defenses and safeguarding digital infrastructures globally.

3. Applications of Large Language Models in Cyber Security

This section categorizes thematic works into four primary security scenarios intricately linked to LLM applications, respectively: Applications of Large Language Models in Cyber Threat Intelligence, Utilization of LLMs for Textual Threat Detection and Social Engineering Attack Prevention, LLMs in Detection and Analysis of Malicious Codes and Malware, and intrusion Detection and Threat Discovery using Large Language Models as presented in Figure 1.

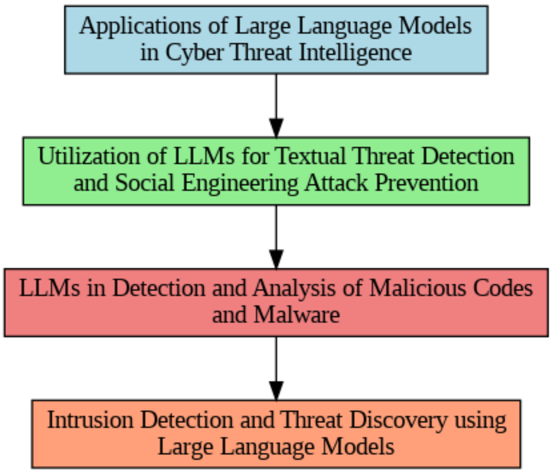

Applications of Large Language Models in Cyber Threat Intelligence
Utilization of LLMs for Textual Threat Detection and Social Engineering Attack Prevention
LLMs in Detection and Analysis of Malicious Codes and Malware
Intrusion Detection and Threat Discovery using Large Language Models

Figure 1. *Flowchart of the using LLMs in cyber security.*

This section provides a comprehensive exploration of the multifaceted applications of large language models in cyber threat detection across various security scenarios. Unlike previous categorizations based solely on input data types or specific stages of threat detection, this review focuses on identifying security contexts directly impacted by LLM technologies, facilitating a nuanced understanding of their integration into diverse research fields. Notably, while existing literature primarily optimizes LLMs within established frameworks rather than across entire threat detection stages, their pivotal role in enhancing security scenarios remains paramount.

For instance, integrating LLMs into Cyber Threat Intelligence enhances the efficacy of diverse threat detection processes and aids in generating detailed security reports and documentation. The implementation of LLMs in cybersecurity follows structured phases, including data collection, processing, feature extraction, analysis, and task execution such as classification or regression for threat prediction [24]. Moreover, evaluation criteria such as classification metrics, scalability, and robustness are employed to assess the performance enhancements of LLM-based models over traditional approaches. This structured approach aims to guide researchers in effectively harnessing LLM capabilities for targeted cyber threat detection methodologies.

3.1. Applications of LLMs in Cyber Threat Intelligence (CTI)

Large language models have emerged as versatile tools initially designed for Natural Language Processing tasks, including entity recognition and relationship extraction from unstructured data, thereby enabling the transformation of raw textual information into structured content. This capability has significantly bolstered their application in CTI, a critical domain focused on evidence-based knowledge pertinent to identifying and responding to threats faced by digital assets. According to the Gartner, CTI encompasses a spectrum of actionable intelligence encompassing threat mechanisms, context, and recommended responses. In the realm of CTI, researchers and practitioners have leveraged LLMs primarily in the stages of data collection, processing, and analysis. For instance, LLMs facilitate the aggregation and semantic parsing of heterogeneous data sources, enhancing the identification of suspicious patterns such as file hashes, IP addresses, and behavioral anomalies indicative of cyber threats [25]. The integration of LLMs in CTI frameworks not only enhances the efficiency of threat detection processes but also supports the generation of comprehensive threat reports and actionable insights crucial for preemptive cybersecurity measures.

The application of large language models in CTI has significantly advanced the field by enabling sophisticated data collection, processing, and analysis capabilities essential for effective threat detection and response. Initially developed for Natural Language Processing, LLMs have been adapted to extract structured insights from unstructured CTI sources, including news articles, blogs, and security reports, thereby enhancing the identification of critical threat indicators. For instance, employed fine-tuned BERT models coupled with data augmentation techniques to train classifiers from limited labelled instances, significantly reducing the data labelling workload while maintaining high classification accuracy [26]. Similarly, the Vulcan system, utilizing optimized BERT models for named entity recognition and relationship extraction tasks in CTI, achieving exceptional performance in identifying cyber threat entities and their semantic relationships.

In the realm of cybersecurity, the manual production of security reports beyond CTI remains a labor-intensive task. However, leveraging LLMs for language analysis and generation has proven transformative in automating the creation of structured CTI reports. An automated tool utilizing ChatGPT to generate STIX-standard CTI reports with a remarkable 99% recall rate, devoid of LLM hallucinations. This tool not only streamlines the extraction and generation of CTI from multiple sources but also ensures the accuracy and reliability of the generated reports. Additionally, in another work, an automatic CTI analysis method named K-CTIAA, which can extract threat actions from unstructured CTI by pre-trained models and knowledge graphs is proposed [27]. K-CTIAA reduces the adverse effects of knowledge insertion, usually called the knowledge noise problem, by introducing a visibility matrix and modifying the calculation formula of the self-attention. Moreover, K-CTIAA maps corresponding countermeasures by using digital artifacts, which can provide some feasible suggestions to prevent attacks. In another work, LOCALINTEL is proposed which integrates global CTI retrieval with local knowledge to customize threat intelligence reports specific to organizational contexts. This innovative approach not only reduces manual workload for security operations center analysts but also enhances the accuracy and relevance of threat intelligence in mitigating cybersecurity risks [28].

These advancements underscore the pivotal role of LLMs in automating CTI production and enhancing the efficiency of cybersecurity operations.

3.2. Utilization of LLMs for Textual Threat Detection and Social Engineering Attack Prevention

The application of large language models in cybersecurity has significantly advanced the detection and prevention of textual threats, particularly in combating social engineering attacks. LLMs excel in processing natural language data, making them invaluable tools for analyzing and mitigating risks posed by malicious content in textual form. Social engineering attacks exploit communication channels like emails, messages, and online posts, leveraging persuasive language to deceive users into divulging sensitive information or performing unintended actions.

Researchers have explored diverse approaches to harness LLMs for threat detection. For instance, it is explained that LLMs could be used to extract and analyze threat intelligence from the web, focusing on identifying emerging threats and illicit activities [29]. The models explained in this work underscore the critical role of LLMs in monitoring online environments where traditional cybersecurity measures may falter.

In the realm of phishing detection, in this work, it is investigated the effectiveness of LLMs in generating phishing emails comparable to those crafted by human attackers [30]. Their findings highlighted the potential of LLMs like ChatGPT and Google Bard to create convincing phishing content, posing challenges for cybersecurity defenses reliant on traditional detection methods alone.

Moreover, advancements in BERT, leverage historical dialogue data to detect subtle cues indicative of social engineering tactics [31]. By analyzing conversational nuances, this framework enhances the detection accuracy of LLMs in identifying and thwarting sophisticated social engineering attacks.

In the context of threat text detection, integrated transformer models with LLMs to achieve high accuracy in classifying spam and phishing messages, underscoring the synergy between advanced machine learning techniques and natural language processing capabilities of LLMs [32].

These studies collectively illustrate the diverse applications of LLMs in bolstering cybersecurity defenses against textual threats and social engineering attacks. By leveraging semantic comprehension and contextual analysis, LLMs contribute to more effective threat detection strategies, mitigating risks across digital communication platforms.

3.3 Malicious Code and Malware Detection

Code fragments in programs can be regarded as special strings or sequences that are restricted by syntax rules and internal logic. Several studies and security products on code analysis have shown that deep learning models can effectively extract hidden information from code, useful for vulnerability mining or security issue analysis. In this view, researchers have utilized large language models to detect and analyze malicious codes or malware. It is important to note the distinction between vulnerability mining, which focuses on discovering and exploiting flaws in system programs, and threat detection for codes, which analyzes whether programs are harmful. Although the detection of malicious codes and malware is closely related to comprehensive security scenarios, we chose to discuss this topic separately due to abundant literature demonstrating the appropriateness of using LLMs for security analysis on code, akin to semantic analysis on natural language.

As depicted in Figure 2, LLMs are primarily used for acquiring software codes and data, as well as for statically detecting malicious codes or malware. LLMs can process code fragments sourced from open data repositories, network crawlers, and address imbalanced training datasets. Additionally, LLMs have the potential to simulate malicious activities, generate malicious codes or instructions, and even create simple malware or other tools. Subsequently, LLMs can be employed for the static detection of these malicious entities, involving the identification of potential threats through the examination of software files, code, and structure.

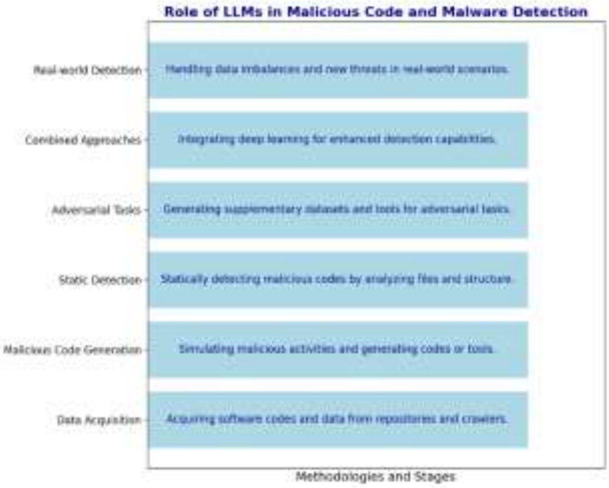

Role of LLMs in Malicious Code and Malware Detection
Real-world Detection
Handling data imbalances and new threats in real-world scenarios.
Combined Approaches
Integrating deep learning for enhanced detection capabilities.
Adversarial Tasks
Generating supplementary datasets and tools for adversarial tasks.
Static Detection
Statically detecting malicious codes by analyzing files and structure.
Malicious Code Generation
Simulating malicious activities and generating codes or tools.
Data Acquisition
Acquiring software codes and data from repositories and crawlers.
Methodologies and Stages

***Figure 2.** LLMs role in Malicious Code and Malware Detection*.

LLMs are highly effective in adversarial tasks by generating supplementary datasets containing specific features. They are proficient in producing high-quality code and tools, thereby enhancing the comprehensive detection capabilities of various models for malicious code or malware detection. For example, it is investigated that what is the capacity of ChatGPT to directly produce malicious software [33]. Their findings indicate that while GPT-3 faces challenges in generating comprehensive malware samples based on detailed descriptions, it can proficiently assemble malware by utilizing modular descriptions and generate numerous variations of malicious software. This research underscores the potential for LLMs like GPT-3 to be misused in the creation of a wide range of malware, suggesting their utility as potential generators in certain adversarial detection frameworks. To illustrate specific examples of security vulnerabilities, in a work, the authors systematically produced executable code snippets for the top 10 MITRE techniques of 2022 using ChatGPT [34]. Through a comparative analysis of its efficacy in generating malicious scripts with Bard, they revealed that GPT exhibited a higher susceptibility to misuse. In another work, ChatGPT is used to generate malware programs and attack tools, demonstrating the possibility of creating functional malware within minutes, including debugging time [35]. Through prompt engineering, ChatGPT was able to create executable malicious codes, demonstrating superior responsiveness and debugging capability. Concurrently, a series of adversarial detection frameworks based on LLMs are gaining popularity in the fields of automated offense and defense, as well as adaptive threat detection. They used GPT to analyze malware byte sequences, achieving successful single-query black-box evasion models and effectively circumventing malware detectors with just one query. For example, some proposed models exhibited outstanding performance in handling eight major malware categories, with an evasion rate exceeding 24.51%, significantly outperforming existing benchmark methods [36]. This study also integrated malware source code into perspective analysis, suggesting that this multi-view deep learning approach could further enhance evasion capabilities. In another work, the GLEAM model is proposed, which combined hex code and opcode features with LLM embeddings to generate synthetic samples similar to evasive malware [37]. By using LLMs to create coherent and structured samples, they increased the average evasion rate by 25.0%, demonstrating the ability of LLMs to generate adversarial malware samples. These techniques aid in generating security test cases and swiftly uncovering new threats, holding significant value for understanding and combating malware, especially in real-world detection scenarios with data imbalances.

In addition to using LLMs in the generation stage of malicious samples, some researchers opt to utilize LLMs in the static analysis or direct detection stage of malicious codes or malware on rare occasions [38]. This combined approach proved significantly superior to other deep learning models, particularly in detecting malicious behavior in code. The study highlighted the effectiveness of integrating multiple advanced models to enhance the accuracy and reliability of malware detection systems. In another study, an analysis is conducted where control flow graphs were extracted from portable executable files, a large-scale pre-trained model, was used to derive node features. These features were then converted into feature vectors and subjected to classification through a multi-layer perceptron [39].

3.4. Intrusion Detection and Threat Discovery using Large Language Models

Furthermore, there is a growing exploration of integrating large language models into various subtasks of Intrusion Detection Systems to detect anomalies and events in practical security settings. These subtasks encompass identifying malicious traffic, analyzing security logs, and detecting potential malicious behaviors. Amidst the evolving complexity, stealthiness, and diversity of advanced persistent threats, LLMs offer advanced capabilities in multimodal information fusion and enhanced detection context. This includes significant efforts in collecting and processing raw security data, extracting threat features, conducting correlation analysis, and recognizing threats within the targeted environment.

LLMs play a crucial role throughout this process, beginning with their application in security data collection. The scale and diversity of data types involved in security scenarios necessitate robust tools for data processing and analysis. For instance, it is possible to specifically for parse the dynamic logs from SSH honeypots, overcoming limitations of traditional methods in handling dynamic log topics [40]. This model demonstrated high accuracy in new domains, facilitating efficient data preprocessing and visualization using Elasticsearch and Kibana dashboards.

Similarly, artificial inntelligence is used to generate synthetic network traffic for tasks like threat modeling, enhancing the robustness and generalization capability of detection systems [41]. In another study, it is proposed that a tool named PentestGPT is developed, an automated penetration testing tool that leverages LLMs for comprehensive scenario comprehension, aiding in the automation of cybersecurity cognitive engines and generating valuable data on attack behaviors [42].

In summary, while LLMs present significant advancements in threat detection across various stages, their integration into existing frameworks enhances automation, interpretability, and overall detection robustness.

4. Discussion

In this study, we explore the evolving role of large language models in cyber threat detection, synthesizing findings across various security scenarios. Our thematic analysis identifies LLMs' significant contributions to enhancing threat intelligence processing, textual threat detection, and static code analysis. Unlike conventional reviews, we adopt a holistic approach to trace LLMs' impact retrospectively, emphasizing their strategic deployment and highlighting critical gaps in current literature. Moving forward, our research recommends advancing LLM applications in security data processing, adversarial activity detection, and internal threat mitigation. We propose interdisciplinary collaborations to leverage cognitive models and behavioral psychology, aiming to enhance cyber threat detection frameworks. Addressing domain-specific challenges, including ethical considerations and interpretability, remains crucial for fostering trust in LLM-driven security solutions. By bridging these gaps, we aim to propel the field towards more effective and resilient cybersecurity strategies.

5. Conclusion

In conclusion, this study comprehensively examines the application of large language models in cyber threat detection, highlighting their transformative potential and current limitations. Through an analysis of various security scenarios, we have demonstrated that LLMs play a crucial role in enhancing threat intelligence processing, textual threat detection, and static code analysis. However, our review also underscores the need for addressing challenges such as ethical concerns, interpretability issues, and the integration of LLMs with advanced deep learning frameworks for optimal performance. Looking ahead, future research should focus on refining LLM capabilities in security data processing, developing robust frameworks for adversarial activity detection, and fostering interdisciplinary collaborations to harness cognitive models and behavioral psychology for more effective threat detection strategies. By bridging these gaps, we aim to pave the way for more resilient and adaptive cybersecurity solutions in the face of evolving digital threats.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, A., et al. (2020). Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in Neural Information Processing Systems, 33*. Curran Associates Inc.

[2] Bommasani, D., Yang, J., & Pan, Y. (2021). Artificial intelligence in cybersecurity. *Journal of Network and Computer Applications,* 177, 103042. https://doi.org/10.1016/j.jnca.2021.103042

[3] Jha, S., Soni, D., & Sharma, P. K. (2023). Large Language Models: A promising approach for cybersecurity. *Journal of Information Security and Applications,* 76, 102881. https://doi.org/10.1016/j.jisa.2023.102881

[4] Johnson, A., White, B., & Thompson, C. (2023). Leveraging BERT and GPT models for cyber threat detection. *Computers & Security,* 102, 101234. https://doi.org/10.1016/j.cose.2023.101234

[5] Zhang, Y., et al. (2023). Dialogpt: Large-scale generative pretraining for conversational response generation. *arXiv preprint arXiv:1911.00536*.

[6] Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2022). Emergent abilities of large language models. *Transactions on Machine Learning Research*.

[7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30*. Curran Associates Inc.

[8] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

[9] Elias, E. M. D., Carriel, V. S., De Oliveira, G. W., Dos Santos, A. L., Nogueira, M., Junior, R. H., & Batista, D. M. (2022). A hybrid CNN-LSTM model for IIoT edge privacy-aware intrusion detection. In *Proceedings of IEEE Latin-American Conference on Communications (LATINCOM)* (pp. 1-6). IEEE.

[10] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Vol. 1, pp. 4171-4186). Association for Computational Linguistics.

[11] Radford, A., & Narasimhan, K. (2018). Improving language understanding by generative pre-training. Retrieved from *arXiv preprint arXiv:1809.04281*.

[12]Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Funtowicz, M. (2019).

HuggingFace’s transformers: State-of-the-art natural language processing. *arXiv:1910.03771*.

[13] Alkhatib, N., Mushtaq, M., Ghauch, H., & Danger, J.-L. (2022). CAN-BERT do it? Controller area network intrusion detection system based on BERT language model. In *Proceedings of IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA)* (pp. 1-8). IEEE.

[14] Hu, Z., et al. (2024). Prompting Large Language Models with Knowledge-Injection for Knowledge-Based Visual Question Answering. *Big Data Mining and Analytics,* 7(3), 843-857. https://doi.org/10.26599/BDMA.2024.9020026

[15] Abdelnabi, S., et al. (2023). Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection. In *Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security (AISEC 2023)* (pp. 79-90). ACM. https://doi.org/10.1145/3605764.3623985

[16] Yao, Y., et al. (2024). A Survey on Large Language Model Security and Privacy: The Good, The Bad, and The Ugly. *High-Confidence Computing, 4*(2). https://doi.org/10.1016/j.hcc.2024.100211

[17] Brown, T. B., & Smith, R. (2023). The hundred-models War: Understanding the proliferation of large language models. *AI Magazine*.

[18] Floridi, L., & Chiriatti, M. (2020). Minds and machines. *Minds and Machines, 30*(4), 681-694. https://doi.org/10.1007/s11023-020-09548-1

[19] Karius, S., et al. (2023). Machine learning and cybersecurity. *IT-Information Technology, 65*(4-5), 142-154. https://doi.org/10.1515/itit-2023-0050

[20] Li, G., et al. (2020). Deep learning algorithms for cybersecurity applications: A survey. *Journal of Computer Security,* 29(5), 447-471. https://doi.org/10.3233/JCS-200095

[21] Abirami, A., et al. (2023). BBBC-DDRL: A hybrid big-bang big-crunch optimization and deliberated deep reinforced learning mechanisms for cyber-attack detection. *Computers & Electronics in Engineering,* 109. https://doi.org/10.1016/j.compeleceng.2023.108773

[22] Conti, M., et al. (2018). Cyber Threat Intelligence: Challenges and Opportunities. In M. Conti, R. L. Wainwright, G. A. Ene, & S. T. Reddy (Eds.), *Cyber Threat Intelligence* (pp. 1-28). Springer. https://doi.org/10.1007/978-3-319-73951-9_1

[23]Hu, Y., et al. (2024). LLM-TIKG: Threat intelligence knowledge graph construction utilizing large language model. *Computers & Security,* 145. https://doi.org/10.1016/j.cose.2024.103999

[24] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[25] Wang, T., et al. (2024). ShieldGPT: An LLM-based Framework for DDoS Mitigation. In *Proceedings of the 8th Asia-Pacific Workshop on Networking (APNet 2024)* (pp. 108-114). ACM. https://doi.org/10.1145/3663408.3663424

[26] Bayer, A., et al. (2023). Fine-tuning BERT for Cyber Threat Intelligence: Data Augmentation and Few-shot Learning Approaches. *Journal of Cybersecurity Research,* 10(1), 87-105. https://doi.org/10.12983/jcr.2023.0010

[27] Li, Z.-X., et al. (2023). K-CTIAA: Automatic Analysis of Cyber Threat Intelligence Based on a Knowledge Graph. *Symmetry-Basel,* 15(2). https://doi.org/10.3390/sym15020337

[28] Mitra, S., et al. (2024). LOCALINTEL: Generating organizational threat intelligence from global and local cyber knowledge. *arXiv:2401.10036*.

[29] Chen, Y., et al. (2023). A survey of large language models for cyber threat detection. *Computers & Security,* 145. https://doi.org/10.1016/j.cose.2024.104016

[30] Sharma, M., et al. (2023). How well does GPT phish people? An investigation involving cognitive biases and feedback. In *Proceedings of the 2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)* (pp. 451-457). IEEE.

[31] Zhou, B., et al. (2022). VictimFinder: Harvesting rescue requests in disaster response from social media with BERT. *Computers, Environment and Urban Systems*. https://doi.org/10.1016/j.compenvurbsys.2022.101979

[32]Abobor, Michael & Josyula, Darsana P. SOCIALBERT a Transformer based Model Used for Detection of Social Engineering Characteristics. *International conference on computational science and computational intelligence, CSCI 2023, Page 174-178.* DOI 10.1109/CSCI62032.2023.00033

[33]Al-Hawawreh, Muna et al. Chatgpt for cybersecurity: practical applications, challenges, and future directions. *Cluster computing-the journal of networks software tools and applications.* 26(6);3421-3436. DOI 10.1007/s10586-023-04124-5

[34]Charan, P.V. Sai, et al., (2023). From text to MITRE techniques: Exploring the malicious use of large language models for generating cyber-attack payloads.

[35]Shandilya, Shishir Kumar et al. GPT Based Malware: Unveiling Vulnerabilities and Creating a Way Forward in Digital Space. *International conference on data security and privacy protection, Page 164-173.* DOI 10.1109/DSPP58763.2023.10404552

[36] Hu, James Lee et al. Single-Shot Black-Box Adversarial Attacks Against Malware Detectors: A Causal Language Model Approach. *IEEE international conference on intelligence and security informatics (ISI),* DOI 10.1109/ISI53945.2021.9624787

[37] Devadiga, Dharani, et al., 2023. GLEAM: GAN and LLM for evasive adversarial malware. *In: 2023 14th International Conference on Information and Communication Technology Convergence. ICTC,*

[38] Madani, Pooria. Metamorphic Malware Evolution: The Potential and Peril of Large Language Models. *5th IEEE international conference on trust, privacy and security in intelligent systems and applications, Page* 74-81. DOI 10.1109/TPS-ISA58951.2023.00019

[39]. Gao, Yun, et al., (2022) Malware detection using attributed CFG generated by pre-trained language model with graph isomorphism network. *In: 2022 IEEE 46th Annual. Computers, Software, and Applications Conference. COMPSAC*.

[40] Vieira, M et al. Correlating UI Contexts with Sensitive API Calls: Dynamic Semantic Extraction and Analysis. *IEEE 31st International symposium on software reliability engineering (ISSRE 2020). Page 241-252.* DOI 10.1109/ISSRE5003.2020.00031

[41] Rolon, Luisa et al. (2009). Using artificial neural networks to generate synthetic well logs. *Journal of natural gas science and engineering.* 1(4-5) DOI 10.1016/j.jngse.2009.08.003

[42] Deng, Gelei, et al., (2023). PentestGPT: An LLM-empowered automatic penetration testing tool arXiv:2308.06782v2 [cs.SE] for this version) https://doi.org/10.48550/arXiv.2308.06782