A Review of Advancements and Applications of Pre-Trained Language Models in Cybersecurity

Conference Paper · April 2024

DOI: 10.1109/ISDFS60797.2024.10527236

1 author:

Zefang Liu
Georgia Institute of Technology
18 PUBLICATIONS **152** CITATIONS

SEE PROFILE

A Review of Advancements and Applications of Pre-trained Language Models in Cybersecurity

Zefang Liu
School of Computational Science and Engineering
Georgia Institute of Technology
Atlanta, USA
liuzefang@gatech.edu

***Abstract*—In this paper, we delve into the transformative role of pre-trained language models (PLMs) in cybersecurity, offering a comprehensive examination of their deployment across a wide array of cybersecurity tasks. Beginning with an exploration of general PLMs, including advancements and the emergence of domain-specific models tailored for cybersecurity, we provide an insightful overview of the foundational technologies driving these developments. The core of our review focuses on the multifaceted applications of PLMs in cybersecurity, ranging from malware and vulnerability detection to more nuanced areas like log analysis, network traffic analysis, and threat intelligence, among others. We also highlight recent strides in the application of large language models (LLMs), showcasing their growing influence in enhancing cybersecurity measures. By charting the landscape of PLM applications and pointing toward future directions, this work serves as a valuable resource for both the research community and industry practitioners, underlining the critical need for continued innovation and exploration in harnessing PLMs to fortify cybersecurity defenses.**

***Index Terms*—cybersecurity, pre-trained language models, large language models, natural language processing, domain-specific models**

I. Introduction

In the rapidly evolving landscape of digital security, the arms race between cyber defenders and attackers continues to escalate. As cyber threats become more sophisticated and pervasive, traditional security measures struggle to keep pace, necessitating more advanced and adaptive approaches. Pre-trained language models (PLMs), which have revolutionized natural language processing (NLP), offer promising avenues for enhancing cybersecurity strategies. By leveraging vast amounts of data to learn complex patterns of human language, PLMs such as BERT [1] and GPT [2] have demonstrated remarkable capabilities in understanding and generating text, making them invaluable tools in the identification, analysis, and mitigation of cyber threats.

The integration of PLMs into cybersecurity applications presents a unique opportunity to address the limitations of conventional security solutions. These models can process and interpret vast and diverse datasets, enabling them to detect anomalies, recognize malicious intent, and predict potential vulnerabilities with high precision. From malware detection to phishing prevention, PLMs are redefining the boundaries of what is possible in cybersecurity, providing more dynamic, proactive, and intelligent systems capable of defending against the ever-changing threat landscape.

However, the integration of PLMs, and particularly the recent progress involving large language models (LLMs) in cybersecurity, introduces both breakthrough capabilities and new vulnerabilities. These advanced models not only enhance threat detection and response but also raise concerns about their misuse. This paper navigates through the landscape of PLM applications within cybersecurity, underlining their transformative impact while critically examining the dual-edged nature of their deployment. Through this exploration, we aim to provide insights into the opportunities and challenges presented by these technologies, paving the way for future research and the development of more secure, intelligent cybersecurity solutions.

II. Pre-trained Language Models

The emergence of pre-trained language models (PLMs) marks a pivotal advancement in the field of natural language processing (NLP), offering profound implications for cybersecurity. This section delves into the general architecture and functionality of PLMs, followed by a discussion of their specific adaptations for cybersecurity applications.

A. General Pre-trained Language Models

PLMs have revolutionized the field of NLP by their ability to understand, generate, and interpret human language with remarkable accuracy. These models undergo pre-training on extensive corpora of text sourced from the internet, literature, and other diverse mediums, which empowers them with a deep understanding of language intricacies, context, and semantics. The most celebrated PLMs are BERT (Bidirectional Encoder Representations from Transformer) [1] and GPT (Generative Pre-trained Transformer) [2], which leverage sophisticated transformer architectures to achieve unparalleled text processing capabilities. Encoder models, including BERT and its variants (RoBERTa [3], DistilBERT [4], ALBERT [5], DeBERTa [6], etc.) are distinguished for their bidirectional text analysis, offering profound insights into context and meaning in language, making them exceptionally effective for text classification tasks. The decoder models, especially GPT 1-3 [2], [7], [8], ChatGPT [9], and GPT-4 [10], push the boundaries in generating coherent, context-aware text, demonstrating

remarkable proficiency in content creation, summarization, and conversation. Encoder-decoder models, such as BART [11], T5 [12], and Flan-T5 [13], excel in tasks that require generating new sentences based on specific inputs. Moreover, the advent of open-source large language models (LLMs) such as Llama [14], [15], Falcon [16], Vicuna [17], Mistral [18], [19], Zephyr [20], Phi [21], [22] have democratized access to cutting-edge NLP technology, encouraging widespread innovation and application across various domains. These models not only continue to enhance the performance on traditional NLP tasks but also pave the way for new applications, embedding a deeper level of linguistic understanding and generative capabilities within AI systems. This broad spectrum of PLMs, from BERT's contextual comprehension to GPT's generative excellence, illustrates the dynamic and evolving landscape of PLMs in driving forward the frontiers of technology and human-machine interaction.

TABLE I
CYBERSECURITY PRE-TRAINED LANGUAGE MODELS

Model	Arch.	Type	Params.	Vocab.
SecBERT	BERT	MLM	83M	52k
SecRoBERTa	RoBERTa	MLM	83M	52k
SecureBERT	RoBERTa	MLM	125M	50k
SecureBERT+	RoBERTa	MLM	125M	50k
SecureDeBERTa	DeBERTa	MLM	198M	128k
CySecBERT	BERT	MLM	109M	31k
CALM2-7B	Llama	CLM	7B	65k
CALM2-7B-Chat	Llama	CLM	7B	65k
CyberBase-13B	Llama	CLM	13B	32k
WhiteRabbitNeo-7B	Llama	CLM	7B	32k
WhiteRabbitNeo-33B	Llama	CLM	33B	32k
Lily-Cybersecurity-7B	Mistral	CLM	7B	32k

*MLM/CLM: Masked/Causal Language Modeling

B. Cybersecurity Pre-trained Language Models

In the realm of cybersecurity, the specialization of PLMs [23] has given rise to tailored solutions addressing the domain's unique challenges. SecBERT [24] was developed as a BERT model trained on cybersecurity texts from various sources, including advanced persistent threat (APT) notes and malware-related texts. Ameri et al. [25] introduced CyBERT, a cybersecurity claims classifier fine-tuned from BERT, which demonstrated superior accuracy in classifying features of industrial control systems (ICS) devices. Ranade et al. [26] presented another variant of CyBERT, focusing on contextualized embeddings for cybersecurity, showcasing its efficacy in cybersecurity-specific tasks using a fine-tuned BERT model on a cybersecurity corpus. Aghaei et al. [27] proposed SecureBERT, aimed at automating critical cybersecurity tasks through a domain-specific language model trained on a large corpus of cybersecurity text, proving its superiority in tasks requiring a nuanced understanding of cybersecurity language. Bayer et al. [28] developed CySecBERT, a language model fine-tuned for the cybersecurity domain, which was rigorously compared with other models on domain-dependent tasks, showing enhanced performance and a strong capability in preventing catastrophic forgetting. Table I presents a summary of cybersecurity PLMs from Hugging Face [29], including their model names, architectures, types, numbers of parameters, and vocabulary sizes.

Benchmarking cybersecurity PLMs is also a critical step toward understanding their capabilities and limitations in addressing cybersecurity challenges. Liu et al. [30] introduced CyberBench, a pioneering tool for evaluating LLMs in cybersecurity tasks, marking a significant advancement in the domain-specific application of LLMs. Bhatt et al. [31] developed CyberSecEval, a comprehensive benchmark for assessing the cybersecurity robustness of coding assistant LLMs, revealing the necessity of integrating security considerations into the development of advanced models. Liu's SecQA dataset [32], generated from cybersecurity textbooks using GPT-4, aims to evaluate LLMs on security principles, establishing a novel benchmark for assessing model understanding in cybersecurity. Tihanyi et al. [33] introduced CyberMetric, a benchmark dataset designed to compare human and LLM performance in cybersecurity knowledge, highlighting LLMs' superior performance across various cybersecurity aspects. Chi et al.'s PLUE benchmark [34] focuses on evaluating natural language understanding (NLU) technologies across multiple tasks in privacy policy analysis, underscoring the benefits of domain-specific pre-training. Shankar et al. [35] presented PrivacyGLUE, emphasizing the importance of in-domain pre-training for privacy policy understanding and showcasing transformer models' performance improvement through domain specialization. Joyce et al.'s MalDICT [36] offers a comprehensive dataset for malware attribute classification, leveraging a sophisticated tagging tool to enhance malware analysis. Lin et al. [37] proposed a benchmarking framework for DL-based vulnerable function detection, facilitating a unified performance evaluation across various architectures. Chen et al. [38] released DiverseVul, a large dataset for vulnerability detection research, pointing out the challenges and potential of deep learning, particularly the promising direction of using LLMs for vulnerability detection.

III. CYBERSECURITY APPLICATIONS

The transformative impact of pre-trained language models (PLMs) across various domains within cybersecurity underscores a significant shift towards more intelligent and automated cybersecurity measures. Our review, emphasizing recent research in 2019-2024, ensures a comprehensive overview of the most contemporary and impactful advancements in PLM applications within cybersecurity, highlighting their significant role in bolstering defenses across various domains.

A. Malware Detection

Malware detection [39] plays a crucial role in identifying and mitigating threats posed by malicious software. Oak et al. [40] explored Android malware detection through dynamic analysis of application activity sequences using BERT, demonstrating effectiveness despite imbalanced datasets. Yesir and

Soğukpinar [41] subjected API call sequences to optimization processes for malware classification using fastText and BERT, showcasing methodological success on diverse datasets. Rahali and Akhloufi introduced MalBERT [42], employing transformers for static analysis of Android app source codes, highlighting the high performance of transformer-based models in malware detection. The same authors further developed MalBERTv2 [43], enhancing malware identification through code-aware BERT-based models, emphasizing the significance of pre-tokenization and training on source code relevance. Xu et al. [44] proposed Malbert for detecting Windows malware via dynamic analysis, underscoring the importance of pre-training in deep learning models. Alvares and Di Troia [45] trained machine learning models with BERT-generated word embeddings for malware classification, comparing its efficiency against Word2Vec. Kale et al. [46] applied word embedding techniques, including Word2Vec, HMM2Vec, BERT, and ELMo, for malware classification, achieving improved performance and training times. Demirci et al. [47] utilized stacked BiLSTM and GPT-2 in detecting malicious code, proving the efficacy of pre-trained models in enhancing detection. Souani et al. [48] conducted empirical studies on Android malware detection and classification with BERT, validating the reproducibility of previous results and the non-essential role of permissions. Saracino and Simoni [49] employed BERT for graph-based Android malware detection, leveraging API call graphs for a deeper behavioral analysis. Lastly, Liu et al.'s SeMalBERT model [50] employed BERT for semantic-based Windows malware detection, incorporating a hybrid discriminator to enhance recognition of malicious software.

B. Vulnerability Detection

Vulnerability detection [51] in cybersecurity is an evolving field where PLMs are increasingly applied to identify and mitigate security risks in software. Ziems and Wu [52] modeled software vulnerability detection as an NLP problem, applying deep learning models with transfer learning for automated detection. Das et al. [53] introduced V2W-BERT, a transformer-based learning framework for automating the mapping of vulnerabilities to weaknesses, achieving high prediction accuracy through NLP and transfer learning. Shahid and Debar [54] leveraged NLP for determining the severity of security vulnerabilities from descriptions, employing BERT classifiers for each metric of the Common Vulnerability Scoring System (CVSS). Hanif and Maffeis [55] developed VulBERTa, pre-training a RoBERTa model on source code for vulnerability detection, demonstrating its effectiveness across various datasets. Zhu et al. [56] introduced a deep learning framework for identifying vulnerability types at a fine granularity, employing BERT to model code slice features. Kim et al. [57] presented VulDeBERT, a tool that fine-tunes BERT on a vulnerable code dataset, showing superior performance in detecting specific vulnerability types. Sun et al. [58] developed ASSBert, combining active and semi-supervised learning with BERT for smart contract vulnerability detection, addressing the challenge of insufficient labeled data. Shestov et al. [59] focused on fine-tuning LLMs for vulnerability detection, demonstrating improved performance through training optimizations and addressing class imbalance.

C. Intrusion Detection

Intrusion detection systems [60] are critical for identifying and mitigating cyber threats in various networks and applications, leveraging the advancements in PLMs for enhanced detection capabilities. Nam et al. [61] proposed an intrusion detection model for in-vehicle networks using a bi-directional GPT to detect attacks on the Controller Area Network (CAN) bus, highlighting the model's ability to identify subtle pattern changes indicative of an intrusion. Seyyar et al. [62] presented a web intrusion detection system using the BERT model and a convolutional neural network (CNN) for classifying normal and abnormal URLs, illustrating the effectiveness of combining NLP techniques with deep learning for security. Nguyen and Watabe [63] applied the BERT model to improve domain adaptation in network intrusion detection, employing NLP techniques for enhanced efficiency in detecting threats across different domains. Alkhatib et al. [64] developed CAN-BERT, a deep learning-based system to detect cyberattacks on the CAN bus protocol, demonstrating the power of BERT for real-time anomaly detection. Nwafor and Olufowobi [65] proposed CANBERT, a language-based model for detecting attacks in vehicular networks, showcasing the model's high precision in identifying a range of cyber threats. Ullah et al. [66] introduced IDS-INT, combining transformer-based transfer learning with the SMOTE technique and a CNN-LSTM model for addressing data imbalance and complex feature interactions in network traffic. Lei et al. [67] explored RP-Bert, a combination of transfer learning and rules for detecting and classifying network intrusions, addressing challenges in model generalization. Finally, Manocchio et al. [68] introduced FlowTransformer, a framework utilizing transformer models for flow-based network intrusion detection, demonstrating the flexibility and efficiency of transformers in capturing complex network behaviors.

D. Phishing Detection

Phishing detection is a critical aspect of cybersecurity, aimed at identifying fraudulent attempts to obtain sensitive information through deceptive communications. Lee et al. [69] proposed a fine-tuned BERT model named CATBERT, which uses a context-aware network to efficiently learn sophisticated representations for detecting social engineering emails by considering both content and context features from email headers. Bountakas et al. [70] compared the effectiveness of combining NLP methods with machine learning techniques for phishing email detection, highlighting the superiority of Word2Vec with random forest and logistic regression algorithms on different datasets. Haynes et al. [71] developed a lightweight, transformer-based phishing detection algorithm for mobile devices, focusing on distinguishing phishing from legitimate websites using URLs and demonstrating the advantages of using pre-trained transformers for real-time detec-

tion on mobile platforms. Maneriker et al. [72] introduced URLTran, a transformer-based model that significantly improves phishing URL detection performance and robustness against adversarial attacks by leveraging domain-specific pre-training tasks. Jonker et al. [73] investigated various NLP and machine learning solutions for phishing detection, presenting their potential in generating effective classification results. Elsadig et al. [74] presented a BERT-based feature extraction and deep learning method for phishing URL detection, showcasing its high accuracy and efficiency without manual feature extraction. Misra and Rayz [75] adapted PLMs for phishing email detection, achieving near-perfect performance on in-domain data and improvements on out-of-domain emails through classification-based fine-tuning and a priming-based approach. He et al. [76] proposed a tiny-Bert stacking-based phishing website detection model, which uses a stacking algorithm-based classifier for higher accuracy and stability. Trad and Chehab [77] explored the effectiveness of LLMs in phishing detection, comparing prompt-engineering techniques with fine-tuning and highlighting the superiority of the latter for task-specific performance. Chataut et al. [78] assessed the effectiveness of LLMs in phishing email detection, showing their proficiency and potential implications for enhancing email security.

E. Spam Detection

Spam detection [79] is a crucial component in safeguarding digital communication from unwanted, potentially harmful content. Irissappane et al. [80] leveraged GPT-2 for adversarial training to classify spam reviews with limited labeled data, enhancing detection accuracy with synthetic review generation. Cao et al. [81] developed a bilingual, multi-type spam detection model using M-BERT, which proved effective across various spam types and languages, incorporating OCR for image-based spam. Liu et al. [82] explored the transformer model for SMS spam detection, achieving optimal results across multiple datasets and indicating its adaptability to other spam detection contexts. AbdulNabi et al. [83] utilized BERT for spam email detection, demonstrating its effectiveness over traditional deep learning and machine learning models by leveraging word embeddings and attention layers. Rifat et al. [84] demonstrated a BERT-based universal spam detection model for real-time SMS spam classification, emphasizing its high accuracy and effectiveness. Tida and Hsu [85] proposed a universal spam detection model with BERT, trained on multiple datasets to efficiently classify emails in real-time scenarios, highlighting its versatility. Sahmoud et al. [86] employed BERT for a comprehensive spam detection model, showcasing high performance across diverse corpora. Debnath and Kar [87] applied deep learning techniques, including LSTM and BERT, for email spam detection, achieving remarkable accuracy and showcasing the superiority of BERT in text classification tasks. Oswald et al. [88] introduced SpotSpam, an intention analysis-driven SMS spam detection approach using BERT embeddings, focusing on semantic and textual features for improved filtering accuracy.

F. Log Analysis

Log analysis [89] plays an essential role in monitoring and securing computer systems by analyzing and processing system-generated logs to detect anomalies, ensuring system reliability and security. Guo et al. [90] introduced LogBERT, a self-supervised framework based on BERT for log anomaly detection that learns normal log patterns through novel training tasks, showing superior performance over traditional methods. Ott et al. [91] proposed a framework utilizing PLMs like BERT, GPT-2, and XL for robust and transferable anomaly detection in log data, achieving high performance and robustness against semantic changes. Wang et al. [92] developed a method combining BERT and VAE for log sequence anomaly detection through dual feature extraction and contrastive adversarial training, outperforming traditional approaches. Zhang et al. [93] presented LogST, leveraging Sentence-BERT for semantic extraction in log events and improving anomaly detection with a GRU model. Le and Zhang [94] proposed NeuralLog, a novel approach that forgoes log parsing by extracting semantic vectors from raw log messages for anomaly detection using a transformer-based model, achieving high accuracy. Chen and Liao [95] developed BERT-Log, utilizing a PLM to learn log sequence semantics for anomaly detection, showing significant performance improvements. Hu et al. [96] introduced LogADSBERT, based on Sentence-BERT, for extracting semantic behavior characteristics of log events, offering high accuracy and robustness. Lee et al. [97] proposed LAnoBERT, a parser-free method using BERT for system log anomaly detection through masked language modeling, enhancing performance compared to traditional models. Liu et al. [98] introduced LogPrompt, employing LLMs with advanced prompt strategies for interpretable online log analysis, significantly improving performance without in-domain training data. Qi et al. [99] proposed LogGPT, exploring ChatGPT's language interpretation capabilities for log-based anomaly detection and providing insights into prompt-based models' applicability. Han et al. [100] developed another version of LogGPT, employing GPT with a reinforcement learning strategy for targeted anomaly detection enhancement. Ma et al. [101] introduced KnowLog, a knowledge-enhanced pre-trained model for log understanding, addressing domain-specific challenges and achieving state-of-the-art results.

G. Network Traffic Analysis

Network traffic analysis is pivotal in cybersecurity for monitoring, detecting, and analyzing network activities to ensure data security and network integrity. Lin et al. [102] introduced ET-BERT, a novel model for encrypted traffic classification by pre-training transformers on large-scale unlabeled data, demonstrating significant improvements in classification tasks with deep contextualized datagram-level representation. Shi et al. [103] proposed BFCN, combining BERT and CNN to capture both global and local features of encrypted traffic, achieving state-of-the-art performance on encrypted traffic classification tasks. In another contribution, Shi et al. [104] developed TSFN, integrating BERT for global feature capture

and LSTM for time-series feature extraction, significantly enhancing the accuracy of malicious traffic classification. Meng et al. [105] made the first attempt in the network field to introduce NetGPT, a generative pre-trained model for network traffic understanding and generation, addressing the challenges of modeling diverse network traffic and optimizing for various tasks. Kholgh and Kostakos [106] presented PAC-GPT, a framework to generate synthetic network traffic using GPT-3, aiming to overcome the scarcity of realistic datasets in cybersecurity with a novel approach to data generation. Wang et al. [107] proposed NetLM, an architecture leveraging LLMs to understand and manage network traffic, introducing multi-modal representation learning for intelligent network management and control.

H. Shell Command Analysis

Shell command analysis is an essential aspect of cybersecurity, aimed at interpreting and identifying malicious activities through the examination of command-line inputs. Setianto et al. [108] developed GPT-2C, a system leveraging GPT-2 to parse honeypot logs, specifically focusing on illegal Unix shell commands with high accuracy and reasonable execution latency, marking an advancement in honeypot-based intrusion detection. Andrew and Lim [109] explored the mapping of Linux shell commands to the MITRE ATT&CK® framework using NLP techniques, assessing the effectiveness of various models in accurately aligning commands with ATT&CK tactics and techniques through cosine similarity scoring. Liu and Buford [110] implemented anomaly detection in command shell sessions using DistilBERT, adopting both unsupervised and supervised learning methods to effectively identify deviations in Unix shell command sessions, showcasing the potential of transformers in security anomaly detection. Shi et al. [111] introduced ShellGPT, a model trained on a shell-language corpus to understand shell language, employing pre-tokenization and a novel pre-training objective for command representation, significantly improving performance on shell language understanding tasks.

I. Security Policy Analysis

Security policy analysis [112] in cybersecurity focuses on understanding and enforcing policies that govern the acceptable use of network and information systems to protect against threats and ensure data privacy. Shi et al. [113] introduced the Network Policy Conversation Engine (NPCE) to translate natural language policy questions into network queries, aiding in verifying policy enforcement using big data and NLP techniques. Elluri et al. [114] developed a framework utilizing BERT and BiLSTM to automatically compare web service policies against General Data Protection Regulation (GDPR), evaluating compliance through context similarity scoring, enhancing regulatory adherence assessment. Maitra and Rudrapal [115] proposed a question-answering approach using BERT for simplifying the understanding of privacy policies, offering users insights into specific concerns without the need to read entire documents. McIntosh et al. [116] explored the use of GPT-4 for generating cybersecurity GRC policies focused on ransomware attack mitigation, demonstrating the potential of AI-generated policies to surpass human-created ones in certain metrics with adequate human moderation and tailored input prompts.

J. Named-Entity Recognition

Named-entity recognition (NER) [117] is a crucial task for extracting meaningful information from unstructured text, aiding in the analysis of cyber threats, and enhancing threat intelligence. Chen et al. [118] proposed a joint BERT model integrating LSTM, ID-CNNs, and CRF for cybersecurity NER, incorporating software dictionary features to improve software entity recognition, significantly outperforming traditional models. Zhou et al. [119] applied BERT with whole world masking alongside a BiLSTM-CRF architecture to the cybersecurity NER task, achieving superior evaluation score across entity types. He et al. [120] introduced a BERT-BiLSTM-CRF model tailored for the network security domain, especially effective in identifying key entity information in network security texts, outperforming various LSTM and BERT combinations. Evangelatos et al. [121] investigated the application of transformer-based models for NER in Cyber Threat Intelligence, demonstrating their effectiveness in extracting cybersecurity-related named entities from open-source threat intelligence reports. Alam et al. [122] presented CyNER, an open-source Python library combining transformer-based models and heuristics for cybersecurity NER, trained on a diverse corpus for extracting a wide range of malware attack details. Wang et al. [123] developed a novel feature integration and entity boundary detection model utilizing PERT and GARU for enhanced NER in cybersecurity, incorporating graph and recurrent neural network features for improved entity boundary detection. Srivastava et al. [124] conducted a study on the effectiveness of word embeddings for cybersecurity NER, comparing general-purpose and domain-specific embeddings, finding that fine-tuned BERT embeddings yielded the best performance.

K. Threat Intelligence

Cyber Threat Intelligence (CTI) is a crucial component of cybersecurity, providing insights into potential and actual cyber threats to enhance defense mechanisms. Wang et al. [125] developed an efficient model for extracting unstructured threat intelligence, employing BERT in conjunction with a dictionary template and reinforcement learning to improve entity recognition and relationship extraction in CTI, enhancing threat intelligence sharing. Zhou et al. [126] designed CTI View, an NLP-based automation system for analyzing unstructured CTI, utilizing BERT enhanced with a GRU layer to extract entities more effectively, thereby improving adaptability to heterogeneous CTI. Yan et al. [127] proposed a feature-weighted BERT-BiGRU method for analyzing IIoT threat intelligence, leveraging ATT&CK knowledge to classify and weight attack behavior, enhancing the accuracy and efficiency of emergency response. Liu et al. [128] introduced TriCTI, a system that discovers actionable CTI from cybersecurity

reports using NLP and a novel trigger mechanism, effectively portraying relationships between IOCs and campaign stages. Grigorescu et al. [129] created a model to automatically link CVEs to ATT&CK techniques using BERT, achieving promising results in enhancing the linkage between documented vulnerabilities and adversary behaviors. Kuehn et al. [130] introduced ThreatCrawl, a BERT-based focused crawler for the cybersecurity domain, designed to automate the scanning of online portals for CTI by dynamically classifying documents and adapting its crawling path, surpassing current state-of-the-art solutions. Ranade et al. [131] demonstrated the potential risk of data poisoning in cyber-defense systems by generating fake CTI using transformers, highlighting the need for systems to discern authenticity from fake intelligence. Song et al. [132] explored generating fake CTI texts with GPT-Neo to show the risk of disseminating false information, proposing methods to detect unreliable content.

IV. Large Language Models in Cybersecurity

The recent surge in applications of ChatGPT and other large language models (LLMs) within cybersecurity marks a significant evolution in the field, offering both innovative solutions and new challenges. Alawida et al. [133] explored the use of ChatGPT for generating cyberattacks, highlighting the technology's dual capability for both advancing cybersecurity defenses and facilitating malicious activities. Chowdhury et al. [134] focused on the potential risks ChatGPT poses to the CIA triad of cybersecurity, underscoring the importance of addressing its capacity to circumvent security systems. Gupta et al. [135] delved into the impact of generative AI on cybersecurity, emphasizing the ethical and privacy implications alongside the potential for both defensive and offensive cybersecurity applications. Al-Hawawreh et al. [136] investigated practical applications and challenges of ChatGPT in cybersecurity, including its use in penetration testing and the risks it poses for cybercrime. Kaheh et al. [137] introduced Cyber Sentinel, a GPT-4 based conversational agent aimed at streamlining security tasks, highlighting the balance between explainability and actionability in AI-driven cybersecurity tools. Kalla et al. [138] discussed the advantages and risks associated with ChatGPT's integration into cybersecurity, presenting a nuanced view of its potential to both enhance security measures and be exploited for malicious purposes. Alawida et al. [139] and Okey et al. [140] both illuminated the darker possibilities of ChatGPT's exploitation in cyberattacks while also noting the broader community's awareness and concern over such issues. Yao et al.'s survey on LLM security and privacy [141] categorizes the good, bad, and ugly aspects of LLM applications in cybersecurity, revealing a complex landscape of potential enhancements and vulnerabilities. Motlagh et al.'s comprehensive review of LLMs in cybersecurity [142] identifies key areas for further research, emphasizing the need for a balanced approach to harnessing LLM capabilities while mitigating associated risks. Together, these studies encapsulate the dynamic interplay between advancing AI technologies like ChatGPT and the ever-evolving domain of cybersecurity, highlighting the need for ongoing vigilance, ethical consideration, and defensive strategies.

V. Future Directions

In the evolving landscape of cybersecurity, the integration of large language models (LLMs) presents promising avenues for research and innovation. Addressing the susceptibility of these models to adversarial threats necessitates the development of sophisticated defense strategies, underscoring the need for security measures that evolve with technological advancements. Concurrently, establishing ethical guidelines and privacy-preserving practices is paramount to ensure the responsible use of LLMs in cybersecurity, safeguarding user trust and data integrity. The potential synergy between LLMs and emerging technologies heralds a new era of cybersecurity solutions with enhanced detection capabilities and response mechanisms. Moreover, an interdisciplinary approach is crucial for tackling the socio-technical challenges associated with LLM deployment, encompassing regulatory compliance, user education, and bias mitigation. Encouraging open collaboration and the sharing of LLM-generated cybersecurity insights can foster a collective defense strategy, bolstering global cybersecurity resilience. As this review paper highlights, embracing these future directions is key to maximizing the benefits of LLMs in the cybersecurity domain, paving the way for more secure and intelligent cyberdefense systems.

VI. Conclusion

This review has provided a comprehensive overview of the advancements and applications of pre-trained language models (PLMs) within the cybersecurity domain, highlighting the transition from general PLM advancements to their specialized applications in cybersecurity. Through an in-depth examination of various use cases—ranging from malware detection and phishing identification to log analysis and beyond—we have showcased the significant impact these models have on bolstering cybersecurity efforts. Our exploration reveals the dual nature of PLMs as both powerful tools for enhancing cyber defense mechanisms and as complex systems requiring careful consideration regarding ethical use, privacy concerns, and potential vulnerabilities. As we conclude, it's evident that the development and application of PLMs in cybersecurity represent a pivotal area of research, poised to influence the future of cyber defense strategies amidst a landscape filled with technological and ethical challenges. This paper aims to inspire further exploration and innovation within this dynamic field, advocating for balanced approaches to leveraging PLMs for a more secure digital environment.

References

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[2] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever *et al.*, "Improving language understanding by generative pre-training," 2018.

[3] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[4] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.

[5] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," *arXiv preprint arXiv:1909.11942*, 2019.

[6] P. He, X. Liu, J. Gao, and W. Chen, "Deberta: Decoding-enhanced bert with disentangled attention," *arXiv preprint arXiv:2006.03654*, 2020.

[7] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever *et al.*, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

[8] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[9] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray *et al.*, "Training language models to follow instructions with human feedback," *Advances in Neural Information Processing Systems*, vol. 35, pp. 27 730–27 744, 2022.

[10] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[11] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *arXiv preprint arXiv:1910.13461*, 2019.

[12] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *The Journal of Machine Learning Research*, vol. 21, no. 1, pp. 5485–5551, 2020.

[13] H. W. Chung, L. Hou, S. Longpre, B. Zoph, Y. Tay, W. Fedus, Y. Li, X. Wang, M. Dehghani, S. Brahma *et al.*, "Scaling instruction-finetuned language models," *arXiv preprint arXiv:2210.11416*, 2022.

[14] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar *et al.*, "Llama: Open and efficient foundation language models," *arXiv preprint arXiv:2302.13971*, 2023.

[15] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale *et al.*, "Llama 2: Open foundation and fine-tuned chat models," *arXiv preprint arXiv:2307.09288*, 2023.

[16] E. Almazrouei, H. Alobeidli, A. Alshamsi, A. Cappelli, R. Cojocaru, M. Debbah, É. Goffinet, D. Hesslow, J. Launay, Q. Malartic *et al.*, "The falcon series of open language models," *arXiv preprint arXiv:2311.16867*, 2023.

[17] L. Zheng, W.-L. Chiang, Y. Sheng, S. Zhuang, Z. Wu, Y. Zhuang, Z. Lin, Z. Li, D. Li, E. Xing *et al.*, "Judging llm-as-a-judge with mt-bench and chatbot arena," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[18] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. d. l. Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier *et al.*, "Mistral 7b," *arXiv preprint arXiv:2310.06825*, 2023.

[19] A. Q. Jiang, A. Sablayrolles, A. Roux, A. Mensch, B. Savary, C. Bamford, D. S. Chaplot, D. d. l. Casas, E. B. Hanna, F. Bressand *et al.*, "Mixtral of experts," *arXiv preprint arXiv:2401.04088*, 2024.

[20] L. Tunstall, E. Beeching, N. Lambert, N. Rajani, K. Rasul, Y. Belkada, S. Huang, L. von Werra, C. Fourrier, N. Habib *et al.*, "Zephyr: Direct distillation of lm alignment," *arXiv preprint arXiv:2310.16944*, 2023.

[21] S. Gunasekar, Y. Zhang, J. Aneja, C. C. T. Mendes, A. Del Giorno, S. Gopi, M. Javaheripi, P. Kauffmann, G. de Rosa, O. Saarikivi *et al.*, "Textbooks are all you need," *arXiv preprint arXiv:2306.11644*, 2023.

[22] Y. Li, S. Bubeck, R. Eldan, A. Del Giorno, S. Gunasekar, and Y. T. Lee, "Textbooks are all you need ii: phi-1.5 technical report," *arXiv preprint arXiv:2309.05463*, 2023.

[23] X. Zhao, J. Lu, C. Deng, C. Zheng, J. Wang, T. Chowdhury, L. Yun, H. Cui, Z. Xuchao, T. Zhao *et al.*, "Domain specialization as the key to make large language models disruptive: A comprehensive survey," *arXiv preprint arXiv:2305.18703*, 2023.

[24] Secbert: A pretrained bert model for cyber security text, learned cybersecurity knowledge. [Online]. Available: https://github.com/jackaduma/SecBERT

[25] K. Ameri, M. Hempel, H. Sharif, J. Lopez Jr, and K. Perumalla, "Cybert: Cybersecurity claim classification by fine-tuning the bert language model," *Journal of Cybersecurity and Privacy*, vol. 1, no. 4, pp. 615–637, 2021.

[26] P. Ranade, A. Piplai, A. Joshi, and T. Finin, "Cybert: Contextualized embeddings for the cybersecurity domain," in *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 2021, pp. 3334–3342.

[27] E. Aghaei, X. Niu, W. Shadid, and E. Al-Shaer, "Securebert: A domain-specific language model for cybersecurity," in *International Conference on Security and Privacy in Communication Systems*. Springer, 2022, pp. 39–56.

[28] M. Bayer, P. Kuehn, R. Shanehsaz, and C. Reuter, "Cysecbert: A domain-adapted language model for the cybersecurity domain," *arXiv preprint arXiv:2212.02974*, 2022.

[29] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz *et al.*, "Huggingface's transformers: State-of-the-art natural language processing," *arXiv preprint arXiv:1910.03771*, 2019.

[30] Z. Liu, J. Shi, and J. Buford, "Cyberbench: A multi-task benchmark for evaluating large language models in cybersecurity," 2024.

[31] M. Bhatt, S. Chennabasappa, C. Nikolaidis, S. Wan, I. Evtimov, D. Gabi, D. Song, F. Ahmad, C. Aschermann, L. Fontana *et al.*, "Purple llama cyberseceval: A secure coding benchmark for language models," *arXiv preprint arXiv:2312.04724*, 2023.

[32] Z. Liu, "Secqa: A concise question-answering dataset for evaluating large language models in computer security," *arXiv preprint arXiv:2312.15838*, 2023.

[33] N. Tihanyi, M. A. Ferrag, R. Jain, and M. Debbah, "Cybermetric: A benchmark dataset for evaluating large language models knowledge in cybersecurity," *arXiv preprint arXiv:2402.07688*, 2024.

[34] J. Chi, W. U. Ahmad, Y. Tian, and K.-W. Chang, "Plue: Language understanding evaluation benchmark for privacy policies in english," *arXiv preprint arXiv:2212.10011*, 2022.

[35] A. Shankar, A. Waldis, C. Bless, M. Andueza Rodriguez, and L. Mazzola, "Privacyglue: A benchmark dataset for general language understanding in privacy policies," *Applied Sciences*, vol. 13, no. 6, p. 3701, 2023.

[36] R. J. Joyce, E. Raff, C. Nicholas, and J. Holt, "Maldict: Benchmark datasets on malware behaviors, platforms, exploitation, and packers," *arXiv preprint arXiv:2310.11706*, 2023.

[37] G. Lin, W. Xiao, J. Zhang, and Y. Xiang, "Deep learning-based vulnerable function detection: A benchmark," in *Information and Communications Security: 21st International Conference, ICICS 2019, Beijing, China, December 15–17, 2019, Revised Selected Papers 21*. Springer, 2020, pp. 219–232.

[38] Y. Chen, Z. Ding, L. Alowain, X. Chen, and D. Wagner, "Diversevul: A new vulnerable source code dataset for deep learning based vulnerability detection," in *Proceedings of the 26th International Symposium on Research in Attacks, Intrusions and Defenses*, 2023, pp. 654–668.

[39] Ö. A. Aslan and R. Samet, "A comprehensive review on malware detection approaches," *IEEE access*, vol. 8, pp. 6249–6271, 2020.

[40] R. Oak, M. Du, D. Yan, H. Takawale, and I. Amit, "Malware detection on highly imbalanced data through sequence modeling," in *Proceedings of the 12th ACM Workshop on artificial intelligence and security*, 2019, pp. 37–48.

[41] S. Yesir and İ. Soğukpinar, "Malware detection and classification using fasttext and bert," in *2021 9th International Symposium on Digital Forensics and Security (ISDFS)*. IEEE, 2021, pp. 1–6.

[42] A. Rahali and M. A. Akhloufi, "Malbert: Malware detection using bidirectional encoder representations from transformers," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2021, pp. 3226–3231.

[43] ——, "Malbertv2: Code aware bert-based model for malware identification," *Big Data and Cognitive Computing*, vol. 7, no. 2, p. 60, 2023.

[44] Z. Xu, X. Fang, and G. Yang, "Malbert: A novel pre-training method for malware detection," *Computers & Security*, vol. 111, p. 102458, 2021.

[45] J. Alvares and F. D. Troia, "Bert for malware classification," in *Artificial Intelligence for Cybersecurity*. Springer, 2022, pp. 161–181.

[46] A. S. Kale, V. Pandya, F. Di Troia, and M. Stamp, "Malware classification with word2vec, hmm2vec, bert, and elmo," *Journal of Computer Virology and Hacking Techniques*, vol. 19, no. 1, pp. 1–16, 2023.

[47] D. Demırcı, C. Acarturk *et al.*, "Static malware detection using stacked bilstm and gpt-2," *IEEE Access*, vol. 10, pp. 58 488–58 502, 2022.

[48] B. Souani, A. Khanfir, A. Bartel, K. Allix, and Y. Le Traon, "Android malware detection using bert," in *International Conference on Applied Cryptography and Network Security*. Springer, 2022, pp. 575–591.

[49] A. Saracino and M. Simoni, "Graph-based android malware detection and categorization through bert transformer," in *Proceedings of the 18th International Conference on Availability, Reliability and Security*, 2023, pp. 1–7.

[50] J. Liu, Y. Zhao, Y. Feng, Y. Hu, and X. Ma, "Semalbert: Semantic-based malware detection with bidirectional encoder representations from transformers," *Journal of Information Security and Applications*, vol. 80, p. 103690, 2024.

[51] S. Chakraborty, R. Krishna, Y. Ding, and B. Ray, "Deep learning based vulnerability detection: Are we there yet," *IEEE Transactions on Software Engineering*, 2021.

[52] N. Ziems and S. Wu, "Security vulnerability detection using deep learning natural language processing," in *IEEE INFOCOM 2021-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2021, pp. 1–6.

[53] S. S. Das, E. Serra, M. Halappanavar, A. Pothen, and E. Al-Shaer, "V2w-bert: A framework for effective hierarchical multiclass classification of software vulnerabilities," in *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2021, pp. 1–12.

[54] M. R. Shahid and H. Debar, "Cvss-bert: Explainable natural language processing to determine the severity of a computer security vulnerability from its description," in *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2021, pp. 1600–1607.

[55] H. Hanif and S. Maffeis, "Vulberta: Simplified source code pre-training for vulnerability detection," in *2022 International joint conference on neural networks (IJCNN)*. IEEE, 2022, pp. 1–8.

[56] C. Zhu, G. Du, T. Wu, N. Cui, L. Chen, and G. Shi, "Bert-based vulnerability type identification with effective program representation," in *International Conference on Wireless Algorithms, Systems, and Applications*. Springer, 2022, pp. 271–282.

[57] S. Kim, J. Choi, M. E. Ahmed, S. Nepal, and H. Kim, "Vuldebert: A vulnerability detection system using bert," in *2022 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*. IEEE, 2022, pp. 69–74.

[58] X. Sun, L. Tu, J. Zhang, J. Cai, B. Li, and Y. Wang, "Assbert: Active and semi-supervised bert for smart contract vulnerability detection," *Journal of Information Security and Applications*, vol. 73, p. 103423, 2023.

[59] A. Shestov, A. Cheshkov, R. Levichev, R. Mussabayev, P. Zadorozhny, E. Maslov, C. Vadim, and E. Bulychev, "Finetuning large language models for vulnerability detection," *arXiv preprint arXiv:2401.17010*, 2024.

[60] D. Chou and M. Jiang, "A survey on data-driven network intrusion detection," *ACM Computing Surveys (CSUR)*, vol. 54, no. 9, pp. 1–36, 2021.

[61] M. Nam, S. Park, and D. S. Kim, "Intrusion detection method using bi-directional gpt for in-vehicle controller area networks," *IEEE Access*, vol. 9, pp. 124 931–124 944, 2021.

[62] Y. E. Seyyar, A. G. Yavuz, and H. M. Ünver, "Detection of web attacks using the bert model," in *2022 30th Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2022, pp. 1–4.

[63] L. G. Nguyen and K. Watabe, "Flow-based network intrusion detection based on bert masked language model," in *Proceedings of the 3rd International CoNEXT Student Workshop*, 2022, pp. 7–8.

[64] N. Alkhatib, M. Mushtaq, H. Ghauch, and J.-L. Danger, "Can-bert do it? controller area network intrusion detection system based on bert language model," in *2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA)*. IEEE, 2022, pp. 1–8.

[65] E. Nwafor and H. Olufowobi, "Canbert: A language-based intrusion detection model for in-vehicle networks," in *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2022, pp. 294–299.

[66] F. Ullah, S. Ullah, G. Srivastava, and J. C.-W. Lin, "Ids-int: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic," *Digital Communications and Networks*, 2023.

[67] S. Lei, X. Zhang, and J. Yi, "Rp-bert: An approach to detect and classify network intrusions based on a combination of transfer learning and rules," in *Journal of Physics: Conference Series*, vol. 2504, no. 1. IOP Publishing, 2023, p. 012061.

[68] L. D. Manocchio, S. Layeghy, W. W. Lo, G. K. Kulatilleke, M. Sarhan, and M. Portmann, "Flowtransformer: A transformer framework for flow-based network intrusion detection systems," *Expert Systems with Applications*, vol. 241, p. 122564, 2024.

[69] Y. Lee, J. Saxe, and R. Harang, "Catbert: Context-aware tiny bert for detecting social engineering emails," *arXiv preprint arXiv:2010.03484*, 2020.

[70] P. Bountakas, K. Koutroumpouchos, and C. Xenakis, "A comparison of natural language processing and machine learning methods for phishing email detection," in *Proceedings of the 16th International Conference on Availability, Reliability and Security*, 2021, pp. 1–12.

[71] K. Haynes, H. Shirazi, and I. Ray, "Lightweight url-based phishing detection using natural language processing transformers for mobile devices," *Procedia Computer Science*, vol. 191, pp. 127–134, 2021.

[72] P. Maneriker, J. W. Stokes, E. G. Lazo, D. Carutasu, F. Tajaddodianfar, and A. Gururajan, "Urltran: Improving phishing url detection using transformers," in *MILCOM 2021-2021 IEEE Military Communications Conference (MILCOM)*. IEEE, 2021, pp. 197–204.

[73] R. A. A. Jonker, R. Poudel, T. Pedrosa, and R. P. Lopes, "Using natural language processing for phishing detection," in *International Conference on Optimization, Learning Algorithms and Applications*. Springer, 2021, pp. 540–552.

[74] M. Elsadig, A. O. Ibrahim, S. Basheer, M. A. Alohali, S. Alshunaifi, H. Alqahtani, N. Alharbi, and W. Nagmeldin, "Intelligent deep machine learning cyber phishing url detection based on bert features extraction," *Electronics*, vol. 11, no. 22, p. 3647, 2022.

[75] K. Misra and J. T. Rayz, "Lms go phishing: Adapting pre-trained language models to detect phishing emails," in *2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. IEEE, 2022, pp. 135–142.

[76] D. He, X. Lv, S. Zhu, S. Chan, and K.-K. R. Choo, "A method for detecting phishing websites based on tiny-bert stacking," *IEEE Internet of Things Journal*, 2023.

[77] F. Trad and A. Chehab, "Prompt engineering or fine-tuning? a case study on phishing detection with large language models," *Machine Learning and Knowledge Extraction*, vol. 6, no. 1, pp. 367–384, 2024.

[78] R. Chataut, P. K. Gyawali, and Y. Usman, "Can ai keep you safe? a study of large language models for phishing detection," in *2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2024, pp. 0548–0554.

[79] İ. Yurtseven, S. Bagriyanik, and S. Ayvaz, "A review of spam detection in social media," in *2021 6th International Conference on Computer Science and Engineering (UBMK)*. IEEE, 2021, pp. 383–388.

[80] A. A. Irissappane, H. Yu, Y. Shen, A. Agrawal, and G. Stanton, "Leveraging gpt-2 for classifying spam reviews with limited labeled data via adversarial training," *arXiv preprint arXiv:2012.13400*, 2020.

[81] J. Cao and C. Lai, "A bilingual multi-type spam detection model based on m-bert," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.

[82] X. Liu, H. Lu, and A. Nayak, "A spam transformer model for sms spam detection," *IEEE Access*, vol. 9, pp. 80 253–80 263, 2021.

[83] Q. Yaseen *et al.*, "Spam email detection using deep learning techniques," *Procedia Computer Science*, vol. 184, pp. 853–858, 2021.

[84] N. Rifat, M. Ahsan, M. Chowdhury, and R. Gomes, "Bert against social engineering attack: Phishing text detection," in *2022 IEEE International Conference on Electro Information Technology (eIT)*. IEEE, 2022, pp. 1–6.

[85] V. S. Tida and S. Hsu, "Universal spam detection using transfer learning of bert model," *arXiv preprint arXiv:2202.03480*, 2022.

[86] T. Sahmoud and D. M. Mikki, "Spam detection using bert," *arXiv preprint arXiv:2206.02443*, 2022.

[87] K. Debnath and N. Kar, "Email spam detection using deep learning approach," in *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, vol. 1. IEEE, 2022, pp. 37–41.

[88] C. Oswald, S. E. Simon, and A. Bhattacharya, "Spotspam: Intention analysis–driven sms spam detection using bert embeddings," *ACM Transactions on the Web (TWEB)*, vol. 16, no. 3, pp. 1–27, 2022.

[89] M. Landauer, S. Onder, F. Skopik, and M. Wurzenberger, "Deep learning for anomaly detection in log data: A survey," *Machine Learning with Applications*, vol. 12, p. 100470, 2023.

[90] H. Guo, S. Yuan, and X. Wu, "Logbert: Log anomaly detection via bert," in *2021 international joint conference on neural networks (IJCNN)*. IEEE, 2021, pp. 1–8.

[91] H. Ott, J. Bogatinovski, A. Acker, S. Nedelkoski, and O. Kao, "Robust and transferable anomaly detection in log data using pre-trained language models," in *2021 IEEE/ACM international workshop on cloud intelligence (CloudIntelligence)*. IEEE, 2021, pp. 19–24.

[92] Q. Wang, X. Zhang, X. Wang, and Z. Cao, "Log sequence anomaly detection method based on contrastive adversarial training and dual feature extraction," *Entropy*, vol. 24, no. 1, p. 69, 2021.

[93] M. Zhang, J. Chen, J. Liu, J. Wang, R. Shi, and H. Sheng, "Logst: Log semi-supervised anomaly detection based on sentence-bert," in *2022 7th International Conference on Signal and Image Processing (ICSIP)*. IEEE, 2022, pp. 356–361.

[94] V.-H. Le and H. Zhang, "Log-based anomaly detection without log parsing," in *2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 2021, pp. 492–504.

[95] S. Chen and H. Liao, "Bert-log: Anomaly detection for system logs based on pre-trained language model," *Applied Artificial Intelligence*, vol. 36, no. 1, p. 2145642, 2022.

[96] C. Hu, X. Sun, H. Dai, H. Zhang, and H. Liu, "Research on log anomaly detection based on sentence-bert," *Electronics*, vol. 12, no. 17, p. 3580, 2023.

[97] Y. Lee, J. Kim, and P. Kang, "Lanobert: System log anomaly detection based on bert masked language model," *Applied Soft Computing*, vol. 146, p. 110689, 2023.

[98] Y. Liu, S. Tao, W. Meng, J. Wang, W. Ma, Y. Zhao, Y. Chen, H. Yang, Y. Jiang, and X. Chen, "Logprompt: Prompt engineering towards zero-shot and interpretable log analysis," *arXiv preprint arXiv:2308.07610*, 2023.

[99] J. Qi, S. Huang, Z. Luan, C. Fung, H. Yang, and D. Qian, "Loggpt: Exploring chatgpt for log-based anomaly detection," *arXiv preprint arXiv:2309.01189*, 2023.

[100] X. Han, S. Yuan, and M. Trabelsi, "Loggpt: Log anomaly detection via gpt," in *2023 IEEE International Conference on Big Data (BigData)*. IEEE, 2023, pp. 1117–1122.

[101] L. Ma, W. Yang, B. Xu, S. Jiang, B. Fei, J. Liang, M. Zhou, and Y. Xiao, "Knowlog: Knowledge enhanced pre-trained language model for log understanding," in *Proceedings of the 46th IEEE/ACM International Conference on Software Engineering*, 2024, pp. 1–13.

[102] X. Lin, G. Xiong, G. Gou, Z. Li, J. Shi, and J. Yu, "Et-bert: A contextualized datagram representation with pre-training transformers for encrypted traffic classification," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 633–642.

[103] Z. Shi, N. Luktarhan, Y. Song, and G. Tian, "Bfcn: A novel classification method of encrypted traffic based on bert and cnn," *Electronics*, vol. 12, no. 3, p. 516, 2023.

[104] Z. Shi, N. Luktarhan, Y. Song, and H. Yin, "Tsfn: A novel malicious traffic classification method using bert and lstm," *Entropy*, vol. 25, no. 5, p. 821, 2023.

[105] X. Meng, C. Lin, Y. Wang, and Y. Zhang, "Netgpt: Generative pretrained transformer for network traffic," *arXiv preprint arXiv:2304.09513*, 2023.

[106] D. K. Kholgh and P. Kostakos, "Pac-gpt: A novel approach to generating synthetic network traffic with gpt-3," *IEEE Access*, 2023.

[107] J. Wang, L. Zhang, Y. Yang, Z. Zhuang, Q. Qi, H. Sun, L. Lu, J. Feng, and J. Liao, "Network meets chatgpt: Intent autonomous management, control and operation," *Journal of Communications and Information Networks*, vol. 8, no. 3, pp. 239–255, 2023.

[108] F. Setianto, E. Tsani, F. Sadiq, G. Domalis, D. Tsakalidis, and P. Kostakos, "Gpt-2c: A parser for honeypot logs using large pre-trained language models," in *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2021, pp. 649–653.

[109] Y. Andrew, C. Lim, and E. Budiarto, "Mapping linux shell commands to mitre att&ck using nlp-based approach," in *2022 International Conference on Electrical Engineering and Informatics (ICELTICs)*. IEEE, 2022, pp. 37–42.

[110] Z. Liu and J. Buford, "Anomaly detection of command shell sessions based on distilbert: Unsupervised and supervised approaches," *arXiv preprint arXiv:2310.13247*, 2023.

[111] J. Shi, S. Jiang, B. Xu, J. Liang, Y. Xiao, and W. Wang, "Shellgpt: Generative pre-trained transformer model for shell language understanding," in *2023 IEEE 34th International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 2023, pp. 671–682.

[112] R. F. Ali, P. Dominic, S. E. A. Ali, M. Rehman, and A. Sohail, "Information security behavior and information security policy compliance: A systematic literature review for identifying the transformation process from noncompliance to compliance," *Applied Sciences*, vol. 11, no. 8, p. 3383, 2021.

[113] P. Shi, Y. Song, Z. Fei, and J. Griffioen, "Checking network security policy violations via natural language questions," in *2021 International Conference on Computer Communications and Networks (ICCCN)*. IEEE, 2021, pp. 1–9.

[114] L. Elluri, S. S. L. Chukkapalli, K. P. Joshi, T. Finin, and A. Joshi, "A bert based approach to measure web services policies compliance with gdpr," *IEEE Access*, vol. 9, pp. 148 004–148 016, 2021.

[115] S. Maitra and D. Rudrapal, "Understanding the insights of privacy policies using bert," in *Proceedings of Third International Conference on Advances in Computer Engineering and Communication Systems: ICACECS 2022*. Springer, 2023, pp. 637–644.

[116] T. McIntosh, T. Liu, T. Susnjak, H. Alavizadeh, A. Ng, R. Nowrozy, and P. Watters, "Harnessing gpt-4 for generation of cybersecurity grc policies: A focus on ransomware attack mitigation," *Computers & security*, vol. 134, p. 103424, 2023.

[117] C. Gao, X. Zhang, M. Han, and H. Liu, "A review on cyber security named entity recognition," *Frontiers of Information Technology & Electronic Engineering*, vol. 22, no. 9, pp. 1153–1168, 2021.

[118] Y. Chen, J. Ding, D. Li, and Z. Chen, "Joint bert model based cybersecurity named entity recognition," in *2021 The 4th International Conference on Software Engineering and Information Management*, 2021, pp. 236–242.

[119] S. Zhou, J. Liu, X. Zhong, and W. Zhao, "Named entity recognition using bert with whole world masking in cybersecurity domain," in *2021 IEEE 6th International Conference on Big Data Analytics (ICBDA)*. IEEE, 2021, pp. 316–320.

[120] B. He and J. Chen, "Named entity recognition method in network security domain based on bert-bilstm-crf," in *2021 IEEE 21st International Conference on Communication Technology (ICCT)*. IEEE, 2021, pp. 508–512.

[121] P. Evangelatos, C. Iliou, T. Mavropoulos, K. Apostolou, T. Tsikrika, S. Vrochidis, and I. Kompatsiaris, "Named entity recognition in cyber threat intelligence using transformer-based models," in *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*. IEEE, 2021, pp. 348–353.

[122] M. T. Alam, D. Bhusal, Y. Park, and N. Rastogi, "Cyner: A python library for cybersecurity named entity recognition," *arXiv preprint arXiv:2204.05754*, 2022.

[123] X. Wang and J. Liu, "A novel feature integration and entity boundary detection for named entity recognition in cybersecurity," *Knowledge-Based Systems*, vol. 260, p. 110114, 2023.

[124] S. Srivastava, B. Paul, and D. Gupta, "Study of word embeddings for enhanced cyber security named entity recognition," *Procedia Computer Science*, vol. 218, pp. 449–460, 2023.

[125] X. Wang, R. Chen, B. Song, J. Yang, Z. Jiang, X. Zhang, X. Li, and S. Ao, "A method for extracting unstructured threat intelligence based on dictionary template and reinforcement learning," in *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. IEEE, 2021, pp. 262–267.

[126] Y. Zhou, Y. Tang, M. Yi, C. Xi, and H. Lu, "Cti view: Apt threat intelligence analysis system," *Security and Communication Networks*, vol. 2022, pp. 1–15, 2022.

[127] J. Yan, Z. Du, J. Li, S. Yang, J. Li, J. Li *et al.*, "A threat intelligence analysis method based on feature weighting and bert-bigru for industrial internet of things," *Security and Communication Networks*, vol. 2022, 2022.

[128] J. Liu, J. Yan, J. Jiang, Y. He, X. Wang, Z. Jiang, P. Yang, and N. Li, "Tricti: an actionable cyber threat intelligence discovery system via trigger-enhanced neural network," *Cybersecurity*, vol. 5, no. 1, p. 8, 2022.

[129] O. Grigorescu, A. Nica, M. Dascalu, and R. Rughinis, "Cve2att&ck: Bert-based mapping of cves to mitre att&ck techniques," *Algorithms*, vol. 15, no. 9, p. 314, 2022.

[130] P. Kuehn, M. Schmidt, and C. Reuter, "Threatcrawl: A bert-based focused crawler for the cybersecurity domain," *arXiv preprint arXiv:2304.11960*, 2023.

[131] P. Ranade, A. Piplai, S. Mittal, A. Joshi, and T. Finin, "Generating fake cyber threat intelligence using transformer-based models," in *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2021, pp. 1–9.

[132] Z. Song, Y. Tian, J. Zhang, and Y. Hao, "Generating fake cyber threat intelligence using the gpt-neo model," in *2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP)*. IEEE, 2023, pp. 920–924.

[133] M. Alawida, S. Mejri, A. Mehmood, B. Chikhaoui, and O. Isaac Abiodun, "A comprehensive study of chatgpt: advancements, limitations, and ethical considerations in natural language processing and cybersecurity," *Information*, vol. 14, no. 8, p. 462, 2023.

[134] M. M. Chowdhury, N. Rifat, M. Ahsan, S. Latif, R. Gomes, and M. S. Rahman, "Chatgpt: A threat against the cia triad of cyber security," in *2023 IEEE International Conference on Electro Information Technology (eIT)*. IEEE, 2023, pp. 1–6.

[135] M. Gupta, C. Akiri, K. Aryal, E. Parker, and L. Praharaj, "From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy," *IEEE Access*, 2023.

[136] M. Al-Hawawreh, A. Aljuhani, and Y. Jararweh, "Chatgpt for cybersecurity: practical applications, challenges, and future directions," *Cluster Computing*, vol. 26, no. 6, pp. 3421–3436, 2023.

[137] M. Kaheh, D. K. Kholgh, and P. Kostakos, "Cyber sentinel: Exploring conversational agents in streamlining security tasks with gpt-4," *arXiv preprint arXiv:2309.16422*, 2023.

[138] D. Kalla and S. Kuraku, "Advantages, disadvantages and risks associated with chatgpt and ai on cybersecurity," *Journal of Emerging Technologies and Innovative Research*, vol. 10, no. 10, 2023.

[139] M. Alawida, B. Abu Shawar, O. I. Abiodun, A. Mehmood, A. E. Omolara, and A. K. Al Hwaitat, "Unveiling the dark side of chatgpt: Exploring cyberattacks and enhancing user awareness," *Information*, vol. 15, no. 1, p. 27, 2024.

[140] O. D. Okey, E. U. Udo, R. L. Rosa, D. Z. Rodríguez, and J. H. Kleinschmidt, "Investigating chatgpt and cybersecurity: A perspective on topic modeling and sentiment analysis," *Computers & Security*, vol. 135, p. 103476, 2023.

[141] Y. Yao, J. Duan, K. Xu, Y. Cai, E. Sun, and Y. Zhang, "A survey on large language model (llm) security and privacy: The good, the bad, and the ugly," *arXiv preprint arXiv:2312.02003*, 2023.

[142] F. N. Motlagh, M. Hajizadeh, M. Majd, P. Najafi, F. Cheng, and C. Meinel, "Large language models in cybersecurity: State-of-the-art," *arXiv preprint arXiv:2402.00891*, 2024.