Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, No.1, August : 2024

APPLICATIONS OF LARGE LANGUAGE MODELS IN CYBER SECURITY: OPPORTUNITIES, CHALLENGES, AND RISKS

Ramesh Cheripelli Associate Professor Department of Information Technology, Vidya Jyothi Institute of Technology, Hyderabad, Telangana, India.
B Eswar Babu Associate Professor Department of Information Technology, Vidya Jyothi Institute of Technology, Hyderabad, Telangana, India.
D Sravanthi Assistant Professor Department of Information Technology, Vidya Jyothi Institute of Technology, Hyderabad, Telangana, India.
K Mallikarjuna Rao Assistant Professor Department of Information Technology, Vidya Jyothi Institute of Technology, Hyderabad, Telangana, India.

ABSTRACT
There has been a lot of advancement in data-centric applications with Large Language Models (LLMs) as of late. LLMs demonstrated strong comprehension capabilities and the capacity to encode context when trained on large text datasets. The intriguing thing is that Generative Pre-trained Transformers took advantage of this capability to get AI closer to being able to replace humans in data-centric tasks. By harnessing this power, cyber threat anomalies may be identified, incident response can be enhanced, and normal security procedures can be automated. We classify cyber defence sectors including threat intelligence, automation, privacy preservation, awareness and training, vulnerability assessment, network security, and ethical standards and give an account of the latest actions of LLMs in this area. It lays up the groundwork for LLM development starting with Transformers and moving on to Pre-trained Transformers and GPT. The next step is a review of each section's recent works, along with an analysis of their relative merits and shortcomings. The difficulties and potential future orientations of LLMs in the field of cyber security are discussed in a dedicated section.Lastly, potential avenues for further study.

Introduction
The remarkable achievements of LLMs in the latter part of 2022 and the beginning of 2023 can be credited to the ability of conversationally fine-tuned LLMs to function at a level comparable to human conversation while interacting with people in general. Generative autoregressors, such as ChatGPT and LLaMA derivatives, are LLMs that are capable of performing sleight of hand. Due to their exceptional performance in conversational tasks, there is a temptation to utilize them for tasks such as element categorization, named entity extraction, pattern recognition, or translation. Nevertheless, there exist more appropriate models for such tasks that demand fewer parameters, data, and pretraining in order to attain equivalent performance. Although commercial state-of-the-art autoregressive models are highly effective in solving the problem, their application may be hindered by issues such as pricing, data privacy, or regulatory restrictions.[1,4]
LLMs are demonstrating potential in the field of cybersecurity. Given the increasing quantity and complexity of cyber threats, there is a pressing requirement for intelligent systems capable of autonomously identifying weaknesses, analyzing malicious software, and reacting to attacks. Recent research has investigated the use of LLMs in various cybersecurity tasks, referred to as LLM4Security from now on. LLMs have been utilized in the field of software security to identify vulnerabilities from both plain language descriptions and source code. They are also capable of generating security-related code, including patches and exploits. These models have demonstrated a high level of accuracy in detecting code snippets that are susceptible to vulnerabilities and producing efficient patches for commonly occurring types of vulnerabilities. [2,5,8]
In addition to analyzing code, LLMs have also been used to comprehend and examine security artifacts at a higher level, such as security regulations and privacy policies. This helps in categorizing documents and identifying possible breaches. LLMs in the field of network security have proven their

capability to identify and categorize several forms of assaults based on network traffic data, such as DDoS attacks, port scanning, and botnet activity. LLMs are demonstrating potential in the field of malware analysis. They are utilized to categorize malware families by analyzing textual reports and behavioral descriptions. LLMs have also been utilized in the domain of social engineering to identify and protect against phishing assaults by examining email content and recognizing deceitful language patterns. In addition, researchers are investigating the application of LLMs to improve strength and durability of security systems. The wide range of applications showcased here highlights the considerable capacity of LLMs to enhance the efficiency and efficacy of cybersecurity practices. This is achieved through the processing and extraction of valuable insights from extensive unstructured text, the acquisition of patterns from extensive datasets, and the generation of pertinent examples for testing and training objectives.[3,6,9]

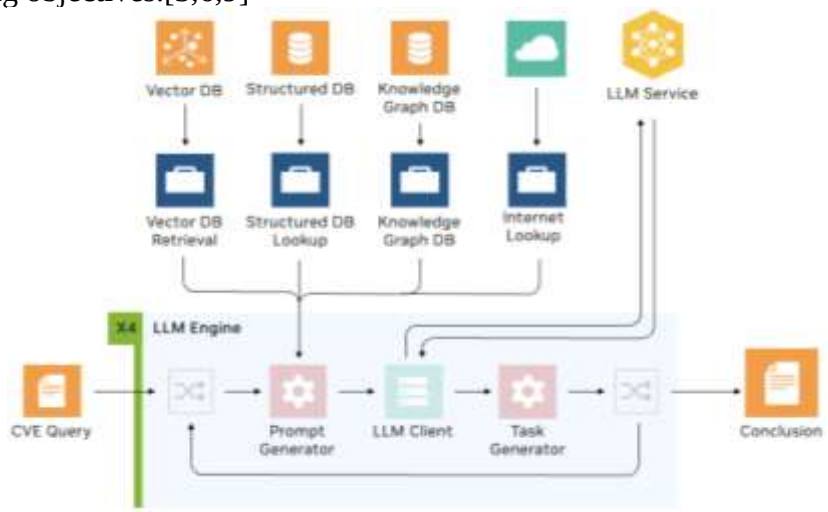

Figure 1: LLM engine that can handle model-generated RAG jobs

RELATED WORKS

Over the course of the past few years, there has been a significant increase in the surface area of the internet, but there has also been an unstoppable surge in the number of new forms of malware that have emerged concurrently with the rapid deployment of innovative online services. In line with the development of malware analysis, the proliferation of creative approaches has also occurred, with the most significant development coming from the advancements made in machine learning.[7,11,13]

The advancements in question are investigated in this research. It is important to note that such approaches frequently adapt to certain settings, which suggests that there may be minimal consistency among the various corporate organizations for the purpose of collecting and keeping tracking information. In the field of cyber threat hunting investigations, MITRE ATT&K serves as the de facto standard for identifying strategies, technologies, and procedures linked to investigating cyber threats. It provides an efficient means of juxtaposing assault samples acquired from various datasets into an efficient threat hunting technique, which is one of the reasons why it plays an increasingly prominent position within CTH investigations. [10,12]

The discovery of botnets is a goal that is supported by research, and certain areas of investigation are centered around this objective. Due to the fact that major botnet cases like as Mirai and Medussa have risen to the forefront and enhanced public understanding of them, this particular topic has received a greater amount of attention for its significance. The analysis of traffic and logs produced by certain networks in order to execute threat hunting by utilizing machine learning models that are tailored to their applications is another example of an innovative activity. [14,17,19]

The cost-effectiveness of this solution in comparison to more traditional approaches, such as constructing and managing in-house Security Operations Center (SOC) teams, is a major contributor

to its widespread adoption. Innovative solutions, such as threat hunting, have resulted in new research programs that are centered on employing its capabilities for discovering zero-day vulnerabilities. This is due to the extraordinary precision that these solutions possess. [15,16,21]

To ensure that it is in line with its primary aim, the suggested algorithm intends to make use of the insights that are gathered from private networks in order to identify any suspicious activities that may indicate the presence of an advanced persistent threat (APT). This is accomplished through the use of methodical searching techniques that are aimed to identify deviations that are indicative of odd behavioral anomalies that may hint at the presence of such dangers. After drawing motivation from current developments in threat hunting through the application of machine learning, we came to the conclusion that it would be beneficial to develop a strategy that is unique to our organization. [18,20,22]

APPLICATION OF LLMS

In this module it investigates use of LLMs in the subject of security related to networks and describes their applications. The activities that it performs include web fuzzing, the detection of anomalies and intrusions, the analysis of cyber threats.

As far as online applications are concerned, security is without a doubt the most important concern. Fuzzing is a technique that can assist operators in identifying additional potential security concerns on web applications. GPTFuzzer was proposed by Liang and colleagues, and it was built on an encoder-decoder architecture. By producing fuzz test cases, it is able to build effective payloads for web application firewalls (WAFs) that are designed to target SQL injection, cross-site scripting, and remote code execution assaults. For the purpose of effectively generating attack payloads and mitigating the local optimum issue, the model is subjected to reinforcement learning fine-tuning and KL-divergence penalty. [23]

In a similar manner, Liu et al. developed a modelon encoder-decoder in order to produce SQL injection detections. This paradigm made it possible for user inputs to be translated into new test cases. On the other hand, Meng et al.'s CHATAFL switches focus to exploiting LLMs for the purpose of creating organized and sequenced effective test inputs for network protocols that do not have machine-readable versions.

One of the most important aspects of network management and security is the detection of network traffic and the detection of intrusions. LLMs have been utilized extensively in network intrusion detection activities, which encompass typical online applications, Internet of Things (IoT) scenarios, and in-vehicle network situations. Not only are LLMs able to understand the features of malicious traffic data and identify anomalies in user-initiated actions, but they are also able to characterize the intentions behind intrusions and anomalous behaviors.Additionally, they are able to provide matching security recommendations and response techniques for the various sorts of attacks that have been found. Using LLMs to extract hierarchical aspects of malicious URLs, Liu et al. suggested a method for identifying malicious URL behavior. [18,24]

Reporting on Cyber Threat Intelligence (CTI) is an essential component of modern risk management techniques, as demonstrated by study that was conducted not too long ago. As a result of the persistent increase in the number of CTI reports, there is an increasing demand for automated technologies that can make the process of report generation easier. When it comes to network threat analysis, the application of LLMs can be broken down into two categories: CTI generation and CTI analysis for decision-making functions. CTI generation can be accomplished in a variety of ways, including the extraction of CTI from network security text material, the development of structured CTI reports from unstructured information, and the generation of CTI from network security entity graphs. The CVEDrill developed by Aghaei et al. is able to provide priority suggestion reports for prospective

cybersecurity risks and then anticipate the impact of those threats. In addition, Moskal et al. investigated the use of ChatGPT to aid or automate the decision-making process for responding to threat behaviors. This experiment demonstrated the potential of LLMs in dealing with straightforward network assault activities. [14,22,23]
There are three stages that make up the basic process of penetration testing: the first is information collecting, the second is payload development, and the third is vulnerability investigation. For the purpose of conducting penetration testing, Temara made use of LLMs to collect information about the target website. This information included the IP address, domain information, vendor technology, SSL/TLS credentials, and various additional details. Sai Charan and colleagues conducted an in-depth analysis of the potential of LLMs to generate harmful payloads for the purpose of penetration testing. The findings of their investigation revealed that ChatGPT is capable of producing payloads that are more complicated and tailored for attackers. LLMs were utilized in the development of an automated Linux privilege escalation advice tool that was by Happe and colleagues. [12,16,21]

4. OPPORTUNITIES AND CHALLENGES

Among the most influential groups in the open-source software security space, OWASP has created a comprehensive list including the most critical flaws connected to LLM-powered apps. Two such kinds of attacks are supply chain vulnerabilities and prompt injection attacks, in which the LLM can be controlled using innovative inputs. Third-party component adoption can cause supply chain risks. Attacks involving quick injection are especially dangerous since they can start from either a compromised data source brought into the prompt or a malicious user input assuming direct access can be obtained. These situations are quite risky both of them. Conversely, the latter might be accomplished by embedding the attack inside a document or website the LLM generates in a retrieval-augmented generation pipeline. Insecure output handling, which can expose backend systems and enable remote code execution, training data poisoning, which may add biases and security risks, and model theft—which may result in the loss of a competitive advantage and sensitive information—are among several other important issues.

In many LLM systems, insecure output handling is a common issue since it affects any solution depending on leveraging model output to initiate actions, execute code, or modify data. This is so since it influences any solution made using model output. Especially considering the concurrent risk of fast injection, building a thorough post-processing pipeline to vet and sanitize LLM outputs against all possible vulnerabilities is a difficult task. Furthermore noted by the Open Web Application Security Project (OWASP) are possible risks include the leaking of private data, assaults rejecting service, insecure design in plugins, too much agency in LLM-based systems, and too great reliance on these models.
Agentic LLM systems deserve particular attention and research since they are vulnerable and prone to all these hazards. Autonomous agents with high-level system rights and comprehensive tool access—like AutoGPT and its more sinister variants—have an even broader attack surface, which can successfully block their responsible use. Should an LLM, agentic or otherwise, be given access to execute arbitrary code in a public-facing application, the cybersecurity of the application is seriously threatened.

The vulnerability database kept by the National Vulnerability Database (CVE2023-29374) claims that LangChain was exposed exactly to this risk until very recently. Furthermore, affected in a similar way are other projects that directly run LLM outputs as code, of which there are surely a huge number now in use. Other applications of agentic LLMs could cause havoc on systems if improper handling of limited rights and surroundings results. This is particularly true in case the system lacks strong security. Apart from the security issues that are naturally present in the design of LLM apps, as mentioned by OWASP, there is another clear category of risks resulting from deliberate use of these applications.

Malicious entities could apply advanced prompt engineering or adversarial fine-tuning of models in order to achieve destructive purposes. Such risks are particularly serious inside the open-source industry since players can more easily bypass built-in safety filters and guardrails of aligned models. There are few mechanisms now available to prohibit such harmful use, and the technical barrier keeping access into this domain is progressively lowering. These weaknesses and risks taken together highlight the need of preventative awareness campaigns, strong security practices, and careful management of open-source development in order to stifle the chance of misuse, data breaches, and other major consequences.

CONCLUSION

Through the enhancement of threat identification, vulnerability assessment, phishing prevention, and security training, Large Language Models have the potential to bring about a huge transformation in the field of the cyber security industry. The implementation of these systems, on the other hand, is fraught with significant dangers, such as risks of misuse, concerns regarding data privacy, bias, and resource intensity. In order to fully utilize the potential of LLMs in the field of cyber security, it is essential to address these problems by implementing robust security measures, explainability, collaborative defensive systems, ethical principles, and scalability efforts. It is imperative that future research and development efforts concentrate on these areas in order to guarantee that LLMs will positively contribute to the creation of a digital landscape that is safer and more secure.

REFERENCES

[1] Abdelrahman Abdallah and Adam Jatowt. 2024. Generator-Retriever-Generator Approach for Open-Domain Question Answering.
[2] Ehsan Aghaei and Ehab Al-Shaer. 2023. CVE-driven Attack Technique Prediction with Semantic Information Extraction and a Domain-specific Language Model.
[3] Ehsan Aghaei, Ehab Al-Shaer, Waseem Shadid, and Xi Niu. 2023. Automated CVE Analysis for Threat Prioritization and Impact Prediction.
[4] Ehsan Aghaei, Xi Niu, Waseem Shadid, and Ehab Al-Shaer. 2023. SecureBERT: A Domain-Specific Language Model for Cybersecurity. In Security and Privacy in Communication Networks, Fengjun Li, Kaitai Liang, Zhiqiang Lin, and Sokratis K. Katsikas (Eds.). Springer Nature Switzerland, Cham, 39–56.
[5] Rio Aguina-Kang, Maxim Gumin, Do Heon Han, Stewart Morris, Seung Jean Yoo, Aditya Ganeshan, R. Kenny Jones, Qiuhong Anna Wei, Kailiang Fu, and Daniel Ritchie. 2024. Open-Universe Indoor Scene Generation using LLM Program Synthesis and Uncurated
[6] Baleegh Ahmad, Benjamin Tan, Ramesh Karri, and Hammond Pearce. 2023. FLAG: Finding Line Anomalies (in code) with Generative AI.
[7] Baleegh Ahmad, Shailja Thakur, Benjamin Tan, Ramesh Karri, and Hammond Pearce. 2023. Fixing Hardware Security Bugs with Large Language Models.
[8] Wasi Uddin Ahmad, Saikat Chakraborty, Baishakhi Ray, and Kai-Wei Chang. 2021. Unified Pre-training for Program Understanding and Generation.
[9] Cheripelli R, New Challenges and its Security, Privacy Aspects on Blockchain Systems,14th International Conference on Advances in Computing, Control, and Telecommunication Technologies, 2023-June,1491-1497
[10] Shunyu Yao et al. React: Synergizing reasoning and acting in language models, 2023
[11] Timo Schick et al. Toolformer: Language models can teach themselves to use tools, 2023.
[12] Tarek Ali and Panos Kostakos. 2023. HuntGPT: Integrating Machine Learning-Based Anomaly Detection and Explainable AI with Large Language Models (LLMs).
[13] Jerry Liu. Llamaindex, 11 2022. If you use this software, please cite it as below.
[14] Natasha Alkhatib, Maria Mushtaq, Hadi Ghauch, and Jean-Luc Danger. 2022. Controller Area Network Intrusion Detection System based on BERT Language Model.

[15] Cheripelli ,New Model to Store and Manage Private Healthcare Records Securely Using Block Chain Technologies, Communications in Computer and Information Science,2022,1550 CCIS,189-201DOI: 10.1007/978-3-031-17181-9_15

[16] Kamel Alrashedy and Abdullah Aljasser. 2024. Can LLMs Patch Security Issues

[17] C. Ramesh,Comparative analysis of applications of identity-based cryptosystem in IoT, Electronic Government, An International Journal, 2017, volume-13,314—323,2017.

[18] Significant Gravitas. Auto-gpt, 2023. An experimental open-source attempt to make GPT-4 fully autonomous.

[19] Matt Bornstein and Rajko Radovanovic. Emerging architectures for llm applications, 2023. Accessed: 2023-08-17.

[20] Cheripelli R et al, Blockchain-Based System for the Secure Transfer of Assets,14th International Conference on Advances in Computing, Control, Telecommunication Technologies,2023, June,885-891

[21] Ross J Anderson and Fabien AP Petitcolas. 1998. On the limits of steganography. IEEE Journal on selected areas in communications 16, 4 (1998), 474–481.

[22] Shishir G. Patil, Tianjun Zhang, Xin Wang, and Joseph E. Gonzalez. Gorilla: Large language model connected with massive apis, 2023.

[23] M Anon. 2022. National vulnerability database. https://www.nist.gov/programs-projects/national-vulnerabilitydatabase-nvd.

[24] Jordi Armengol-Estapé, Jackson Woodruff, Chris Cummins, and Michael F. P. O'Boyle. 2024. SLaDe: A Portable Small Language Model Decompiler for Optimized Assembly.