Large Language Models in Cybersecurity: Upcoming AI Trends in 2023-24

Aleena Noor
aleena.noor@cirruslabs.io

***Abstract*—** Artificial Intelligence (AI) has revolutionized various domains, including healthcare and finance, by enabling machines to imitate intelligent human practices. One area of AI research that has garnered significant attention and sparked unprecedented advancements is large language models (LLMs). LLMs, such as ChatGPT, have transformed natural language understanding, generation, and application. These models, built upon the transformer architecture, leverage billions of parameters and extensive training data to recognize, summarize, translate, predict, and generate content. In the field of cybersecurity, LLMs show promising potential for accelerating threat detection, malware detection, attack defense, and vulnerability assessments. However, their integration into cybersecurity practices raises concerns regarding data privacy, adversarial attacks, and potential system restrictions. This paper provides an overview of LLM frameworks, explores their case uses and applications in various sectors, and examines Microsoft's Security Copilot as an example of LLM application in cybersecurity, highlighting the benefits, limitations, and implications of integrating LLMs into existing security measures.

1 INTRODUCTION

Artificial Intelligence (AI) has revolutionized numerous domains, ranging from healthcare to finance, by enabling machines to perform tasks imitating intelligent human practices. In recent years, one particular area of AI research has captured significant attention and sparked unprecedented advancements: large language models (LLMs). These LLMs have changed the trajectory in natural language understanding, generation, and application.

LLMs take place in the machine learning branch of artificial intelligence which utilizes data and algorithms to imitate intelligent human behavior. A LLM is a

type of a deep learning model consisting of a neural network with billions of parameters, trained on distinct large amounts of data using self-supervised learning. Deep learning is a subset of machine learning with a neural network of multiple layers which optimize and refine for accuracy. LLM as a transformer model, can recognize, summarize, translate, predict and generate content based on knowledge from large data sets.

LLM has changed the trajectory of natural language processing in the past few years with models such as ChatGPT. ChatGPT features an LLM capable of generating text on an endless range of topics. Although still developing, organizations across all industries have shown interest in utilizing these models. In specific, LLM shows promising efficiency in the cybersecurity of companies, a specialized area that is rapidly adapting AI. AI researchers find LLMs could help accelerate efforts to detect data breaches and pinpoint organization vulnerabilities in advance of an attack. At its core, LLMs offer threat detection, malware detection, attack defense, and assessments quickly and effectively. These potential benefits can revolutionize cybersecurity practices. However, it is important to note there are concerns when integrating LLM into these practices.

This paper will outline the framework of LLMs and its latest trends and developments in this field, specifically related to cybersecurity. In addition, it will explore potential case uses, challenges, and applications of LLMs across various sectors.

2 LLM FRAMEWORK

Before analyzing the case uses of LLM, it is essential to delve into the frameworks at the basis of LLM. LLMs are built upon transformer architecture, a type of neural network architecture that relies heavily on self-attention mechanisms. The transformer architecture is highly relevant to cybersecurity fields due to its ability to handle large amounts of textual data and capture complex relationships within that data. LLM is an adaptive version of transformer architecture. Thus, by leveraging this architecture LLMs can effectively analyze cybersecurity-related text, such as network logs, security alerts, or threat intelligence reports. Through analyzing the building blocks and general architecture of LLM, the impact of LLMs on cybersecurity can be better understood.

2.1 General Architecture

The transformer architecture is the fundamental building block for all LLMs. It consists of an encoder-decoder structure with stacked layers of self-attention and feed-forward neural networks. The encoder takes input tokens and generates contextualized representations through self-attention, capturing dependencies and relationships between tokens. The decoder utilizes self-attention and an additional masked attention mechanism to attend to previously generated tokens. Positional encoding is incorporated to provide sequential information. The transformer architecture enables LLMs to effectively process and understand language by capturing contextual information, handling long - range dependencies and facilitating parallel processing.

LLMs are typically trained on massive datasets. Such massive datasets are fed into the AI algorithm using unsupervised learning, in which a model is given a dataset without explicit instructions on what to do with it. Through this, the model learns words and the relationships and context behind them. This learning can be altered to better suit the goals for the model.

2.2 Layers of LLM

LLM architecture primarily consists of multiple layers of neural networks such as recurrent layers, feedforward layers, embedding layers, and attention layers. These layers come together to process the input and generate output predictions.

- Recurrent Layers: These layers are commonly used LLMs to capture sequential information in text data. By maintaining a hidden state that carries information from previous tokens they enable the model to process sequences of tokens. These layers can be employed to model the temporal behavior of network traffic or system logs and can aid analyzing malware-related text data such as binary code.
- Feedforward Layers: These layers are essential as they perform mathematical transformations on the input data, mapping it to a higher dimensional space and enabling non-linear transformations.
- Embedding Layers: These layers map the input tokens (words or sub-words) to dense numerical vectors called embeddings. These embeddings capture the semantic and contextual information for the tokens, allowing the model to understand the meaning relationships between them. These layers are

typically initialized with random values and are fine-tuned during training to capture the specific characteristics of the input data.

- Attention Layers: These layers are a crucial component in LLMs, particularly in transformer-based architecture which ties in with cybersecurity applications. These layers allow the model to focus on different parts of the input sequence when making predictions. Thus, it captures dependencies between tokens. These layers can be broken into two subtypes. One is intra-attention, an attention mechanism where the model attends to different parts of the input sequence computing attention weights for each token based on its relationship with other tokens. The other is multi-head attention, a variant in which multiple attention heads operate in parallel attending to different parts of the input sequence simultaneously. Multi-head attention is commonly seen in transformer based LLMs.

2.3 Training Pipeline

- LLMs features the concept of transfer learning in which pertain models on general language tasks are fine-tuned using smaller datasets. This would include fine-tuning the primary model on cybersecurity-specific data, such as network logs and security alerts. Through this, the model can adapt to the domain. As a result of training on cybersecurity data, the model learns to recognize patterns and relationships of cybersecurity tasks. The trained model can be used to significantly enhance scanning and filtering for cyber security vulnerabilities. In a recent report, the Cloud Security Alliance (CSA) demonstrated that OpenAI's Codex API can effectively scan for vulnerabilities in programming languages like C, C#, Java and JavaScript. For example, a scanner could be constructed to identify and flag insecure code patterns in assorted languages, enabling developers to address security issues before they turn into critical security risks.

3 CASE USES

Large language models have allowed organizations to reach remarkable feats in discovery and innovation in a plethora of field-based/technical applications. These deep learning language models have paved the way to the opportunity of progression. This aspect of continuous improvement is vital with the growing use and development of technology. Thus, creating an emphasis on cybersecurity to ensure the protection of varying components of the computer systems

such as networks, data, digital assets, etc. from cyber threats. LLMs assist in cybersecurity measures through their ability to enhance and, in a way, redefine traditional processes towards a visionary future.

3.1 Computer System Based Uses

LLMs provide an edge in the realm of cybersecurity through their analytical scope.

3.1.1 *Threat Detection and Analysis*

LLMs can analyze great volumes of collected data such as security logs, network traffic, etc. to detect patterns. Observed patterns can then be deciphered to be a threat. A potential threat that is prominent in the cyber world is malware. Through large language models, detection and classification of malware by analyzing code snippets, behavior patterns, and indicators of compromise is made easier. They can assist in identifying new and unknown malware strains, enhancing threat detection capabilities. Therefore, the model lessens the overall widespread negative impacts that malware, among other cyber threats, may have on technologies. Microsoft recently introduced Security Copilot, an AI-driven security analysis tool that empowers analysts to rapidly address potential threats. This chatbot can generate PowerPoint slides that concisely summarize security incidents, detail the extent of exposure to active vulnerabilities, and identify the specific accounts implicated in an exploit, all in response to a user's text prompt. The system combines the power of GPT-4 with the security expertise of Microsoft's security-specific model, which is built from fine-tuning Microsoft activity data.

3.1.2 *Vulnerability Assessment and Patch Management*

Large language models can help in vulnerability assessment of programs through inspecting security advisories, patch notes, and system configurations. The tool can work towards identifying weaker components and recommending mitigations in an optimized pursuit to advance and improve the application. Security researcher Patrick Ventuzelo, founder of cybersecurity services platform Fuzzing Labs, recently used GPT-4 to find zero-day vulnerabilities (undiscovered and unpatched flaws) in snippets of code.

3.1.3 *Anomaly Detection*

LLMs can analyze system logs, network traffic, and other data sources to detect anomalies and potential security breaches. They detect unusual behavior, suspicious patterns, and indicators of compromise that may go unnoticed by traditional systems. Moreover, LLMs can also examine behavioral biometrics through user behavioral data. Specific to detect anomalies, this would be done through comparisons in typing patterns, mouse movements, and navigation behavior. By identifying potential account compromise or unauthorized access attempts through log-in activity or user-pattern, an enhanced authentication and fraud detection mechanism is brought to life.

3.2 Human Interaction Based Uses

The generative language models provide great benefit in cybersecurity, ironically, in the aspect of human-oriented counterparts.

3.2.1 *User Awareness Training*

The models can be used to develop interactive and personalized security awareness training programs. Training can strengthen a user's foundation in a deeper understanding of security concepts and offer guidance on safe online practices. A specific way this can happen is through security incident simulations as they can model scenarios of cybersecurity attacks, which then allows for organizations to be better prepared though the practiced incident response exercises. Thus, enabling individuals to be able to evaluate incidents of security threats with improved response capabilities, while also incorporating their refined strategies learned in the simulations.

3.2.2 *Security Policy and Compliance*

LLMs can aid in the development and enforcement of security policies and compliance frameworks. They can provide guidance on industry best practices, regulatory requirements, and security standards, ensuring organizations adhere to necessary security controls. Additionally, the model can generate code in idea generation or being incorporated in its build when creating programs for security purposes.

4 VULNERABILITIES

While LLMs show significant potential in the cybersecurity field, there are certain risks and potential dangers associated with their introduction.

4.1 Data Privacy and Security

LLMs require access to large volumes of data, which in the cybersecurity field are likely to include sensitive or confidential information. The storage, handling, and processing of such data can pose security concerns. Inadequate data protection measures or any breaches in the infrastructure could lead to unauthorized access, data leaks, or misuse of sensitive information, compromising consumer privacy and organizational security.

4.2 Adversarial Attacks

Cyberattacks on the LLM can lead to organizational vulnerability. Attackers can manipulate LLMs to reveal sensitive information by crafting prompts to disclose confidential data. In addition, attackers can mislead the LLM to perform unintended actions through misleading context. If the LLM were to be implemented in the cybersecurity field, it would hold extreme security risk and harmful outcomes.

4.3 Restricted System

LLM systems can make it harder to detect unexpected activity. This AI technology has the potential to emulate real data such as user activity, network traffic, credentials or profiles – thus restricting any type of stretch mechanism that looks for anomalies.

5 APPLYING THE LLM MODEL: MICROSOFT'S SECURITY COPILOT

Just in this past May, Microsoft released their Security Copilot which combines a Open Large Language Model with a security-specific model. It is the first security product to allow defenders to move at the speed and scale of AI. This section will take a deeper dive into this security model.

5.1 General Infrastructure

Microsoft's Security Copilot model features enterprise grade security and a privacy-compliant experience as it runs on Azure's hyperscale infrastructure. Azure architecture runs on a massive collection of servers and networking hardware. It is made up of physical datacenters, arranged into regions, and linked by a large, interconnected network. Azure architecture includes five main pillars: reliability, cost optimization, operational excellence, performance efficiency, and security. These pillars can be prioritized depending on the task. In the Security Copilot. Security is prioritized to achieve optimal model performance.

5.2 The innerworkings of the Copilot

After the Copilot receives a prompt from a security professional, it uses the full power of the security-specific model to deploy skills and queries that maximize the value of the LLM. Through this it works to analyze an organization's security data and provide recommendations to improve the security posture. It uses AI and machine learning to identify security issues and provide insights to better help protect an organization's assets

5.3 Benefits of the Copilot

While the current majority of security measures in place are efficient, AI takes it to the next level. The Security Copilot allows for a learning system with fine-tuned skills that can be catered to the task. Additionally, it can help catch what other approaches may miss and augment analysts' work. It surpasses past measures as it gains in quality of detection, speed of response, and ability to strengthen security posture.

- Simplifies the Complex: In a field where time plays a huge role, the Security Copilot allows defenders to respond within minutes rather than delaying issues for hours or even days. In addition, the Copilot accelerates incident

investigations as it uses natural language process to provide step-by-step guidance and context.

- Catches Hidden Irregularities: It can be easy for previous measures to miss malicious behavior, but with Security Copilot it can catch these behaviors that would otherwise go undetected. The Copilot includes fine-tuning the LLM which allows it to be knowledgeable on previous malicious behavior and unknown ones. With this, the Copilot can detect the attacker's next move and work to quickly solve the incident or provide guidance on how to.
    - More with less: A Security team's capacity will always be limited by the team size and work hours. The Copilot continually learns based on previous incidents and user interactions. Thus, it is able to adjust to changes or new attacks quickly and provide guidance easily.

6 EXPERIMENTING WITH A SECURITY MODEL

The Microsoft model has yet to be released for public use. By utilizing a pre-trained LLM and The National Institute of Standards and Technology Framework (NIST) cybersecurity framework public data set, trials and experiments can be run. This section will delve into the implications of a security based LLM.

6.1 Benefits of the model

The trial model is far from as advanced as the Microsoft Security Copilot, but it was still capable of answering basic security questions on false emails, incidents, and code. The model could provide suggestions/recommendations on potential responses when asked for a prompt related to a security incident by utilizing its knowledge on the concepts of cybersecurity it was given. Additionally, the trial model could detect well-written emails written through AI that could otherwise go undetected and risk a compromise attack.

6.2 Limitations of the model

Although the trial model may not serve as an accurate demo of the Copilot, it still highlights the limitations when LLM takes place in cybersecurity. The model may be capable of performing defender tasks efficiently, but it will take time and lots of training to successfully integrate the Copilot into procedures and processes that already exist within an organization. When taking a look at the concept from the trial model, it is notable that there will be bias based on the dataset provided. The knowledge base would have to be constantly updated in order for the model to stay up to date with the organization. In addition, the trial model has limited understanding of the context given and may not be increasingly efficient compared to human defenders.

7 CONCLUSION

Large language models (LLMs) have emerged as a powerful tool in the field of cybersecurity, showcasing their potential to enhance threat detection and vulnerability assessments. Microsoft's Security Copilot serves as a prominent example of an LLM application in this domain. By leveraging billions of parameters and extensive training data, LLMs like ChatGPT demonstrate impressive natural language understanding and generation capabilities. However, the integration of LLMs into cybersecurity practices must be approached with caution,

considering concerns related to data privacy, adversarial attacks, and system restrictions. As the trial model evolves, further research and development are required to address these challenges and ensure the safe and responsible implementation of LLMs in the cybersecurity landscape. With careful use and continued development, LLMs have the potential to transform the field of cybersecurity.

8 REFERENCES

1. CSA (2023) How ChatGPT Can Be Used in Cybersecurity. In *Proceedings of the Cloud Security Alliance.*
2. Hore, S. (2023). What are Large Language Models (LLMs)? In *Proceedings of Analytic Vidhya.*
3. Lee, A. (2023) What are Large Language Models Used For? In *Proceedings of Nvidia.*
4. Nadella, S. (2023). Introducing Microsoft Security Copilot. In *Proceedings of the 2023 Virtual Secure Security Conference.*