Chapter 12
Challenges and Limitations of Using LLMs in Software Security

Luay Albtosh

https://orcid.org/0009-0009-0338-9123
Capitol Technology University, USA & Houston Community College, USA

ABSTRACT

Large language models (LLMs) have revolutionized various fields, including software security, by enabling sophisticated analysis and automation. However, despite their potential, the application of LLMs in software security is not without challenges. This chapter explores the limitations of LLMs in this domain, focusing on issues such as data bias, model interpretability, scalability, and the potential for adversarial attacks. The chapter also discusses the complexities of integrating LLMs into existing security frameworks and the ethical implications of their use. By understanding these challenges, researchers and practitioners can better navigate the evolving landscape of software security.

INTRODUCTION

The advent of Large Language Models (LLMs) has marked a significant milestone in the field of artificial intelligence, revolutionizing a wide range of applications, including software security. These models, characterized by their ability to understand and generate human-like text, offer unprecedented capabilities in automating security tasks, detecting vulnerabilities, and enhancing threat intelligence. However, alongside

DOI: 10.4018/979-8-3693-9311-6.ch012

these advancements, the integration of LLMs into software security frameworks has revealed several challenges and limitations that require careful consideration.

One of the most prominent challenges is the issue of model interpretability. While LLMs like GPT-3 and BERT have demonstrated remarkable proficiency in understanding and generating text, the underlying decision-making processes of these models remain largely opaque (Aghaei et al., 2022; Ferrag et al., 2023). This lack of transparency poses significant risks, especially in cybersecurity, where understanding the reasoning behind a model's output is crucial for effective threat detection and mitigation.

Data bias is another critical concern. LLMs are trained on vast amounts of text data, which inevitably contain biases reflective of societal prejudices and historical inaccuracies. These biases can be inadvertently learned by the models, leading to skewed or discriminatory outcomes in security-related tasks (Alawida et al., 2023; Gennari et al., 2024). The implications of such biases are particularly severe in cybersecurity, where equitable and accurate analysis is essential for protecting diverse user populations.

Scalability also presents a significant hurdle. The deployment of LLMs in large-scale security operations demands considerable computational resources and infrastructure, which may not be readily available to all organizations (Al-Hawawreh et al., 2023; Ameri et al., 2021). This raises concerns about the accessibility and feasibility of utilizing LLMs for comprehensive cybersecurity strategies, particularly in resource-constrained environments.

Moreover, the potential for adversarial attacks on LLMs has emerged as a growing threat. Adversaries can exploit vulnerabilities in these models to manipulate their outputs, leading to incorrect or harmful security decisions (Jin et al., 2024; Xu et al., 2024). Such attacks highlight the need for robust defense mechanisms to safeguard LLMs and ensure their reliable application in security contexts.

Lastly, the ethical implications of using LLMs in software security cannot be overlooked. The automation of security processes through LLMs raises questions about accountability, privacy, and the potential for misuse (Motlagh et al., 2024; Wright et al., 2012). As these models become more integrated into cybersecurity frameworks, it is imperative to address these ethical considerations to prevent unintended consequences and ensure responsible use.

In this chapter, we will explore these challenges and limitations in greater detail, drawing on recent research and case studies to provide a comprehensive overview of the current state of LLMs in software security. By examining these issues, we aim to offer insights that can guide the development and deployment of LLMs in a manner that maximizes their benefits while mitigating their risks.