Chapter 2
Foundations of Large Language Models in Software Vulnerability Detection

Hewa Majeed Zangana
https://orcid.org/0000-0001-7909-254X
Duhok Polytechnic University, Iraq

Derek Mohammed
Saint Leo University, USA

ABSTRACT

This chapter explores the foundational aspects of large language models (LLMs) and their application in detecting software vulnerabilities. As the complexity of software systems grows, traditional methods of vulnerability detection are often insufficient. LLMs, with their advanced natural language processing capabilities, provide a novel approach to identifying potential security threats in codebases. The chapter delves into the architecture of these models, their training mechanisms, and the challenges they face in the domain of cybersecurity. Additionally, it discusses the ethical implications and future directions for integrating LLMs in automated software vulnerability detection.

DOI: 10.4018/979-8-3693-9311-6.ch002

INTRODUCTION

The rapid evolution of cyber threats has necessitated the development of advanced tools and techniques for detecting and mitigating software vulnerabilities. In recent years, large language models (LLMs) have emerged as a powerful technology in the field of natural language processing (NLP), with applications that extend into cybersecurity. These models, which have been trained on vast amounts of text data, are capable of understanding and generating human-like text, making them suitable for a variety of tasks, including software vulnerability detection.

One of the significant advantages of LLMs in cybersecurity is their ability to analyze large volumes of code and identify potential vulnerabilities that may be overlooked by traditional methods. As demonstrated by Aghaei et al. (2022), domain-specific language models like SecureBERT have been tailored to address cybersecurity challenges, showing that LLMs can be fine-tuned for specific tasks to improve their accuracy and effectiveness. Similarly, the work of Ranade et al. (2021) on CyBERT highlights how contextualized embeddings can enhance the detection of cybersecurity threats.

The application of LLMs in cybersecurity is not without challenges. According to Gennari et al. (2024), evaluating LLMs for cybersecurity tasks requires careful consideration of their performance, biases, and potential ethical implications. Alawida et al. (2023) also emphasize the limitations and ethical concerns surrounding the use of models like ChatGPT in cybersecurity, particularly in terms of privacy and data security. These challenges underscore the need for a comprehensive approach to integrating LLMs into cybersecurity workflows.

Despite these challenges, the potential of LLMs to revolutionize cybersecurity is evident. Ferrag et al. (2023) argue that LLMs can significantly enhance cyber threat detection, providing a new paradigm for analyzing and responding to emerging threats. This is further supported by the work of Omar and Zangana (2024), who explore how LLMs can be leveraged to enhance security posture and efficiency in various cybersecurity applications.

The integration of LLMs into cybersecurity also presents opportunities for innovation. For instance, the use of LLMs for automated penetration testing, as explored by Happe and Cito (2023), demonstrates the potential for these models to assist in identifying vulnerabilities in a proactive manner. Moreover, the research by Jones et al. (2024) on the GPT-2 Enhanced Attack Detection and Defense (GEADD) method illustrates how LLMs can be utilized to detect and defend against zero-day threats.

As the field of cybersecurity continues to evolve, the role of LLMs in vulnerability detection and threat mitigation will likely expand. The work of Nguyen et al. (2024) on 6G security challenges and opportunities illustrates the growing importance of LLMs in future cybersecurity landscapes. Additionally, the studies