Case Study on
"Utilizing LLMs in Cybersecurity for Code Vulnerability Detection"

A Case Study Report Submitted
For
Partial Fulfillment of the Requirements of the
Degree of Bachelor of Engineering

In

COMPUTER ENGINEERING

(Semester VII)
By

Aditya Vyas 9238

Hitesh Sharma 9233

Atharva Pawar 9427

Under the guidance of

Prof. Supriya Kamoji

DEPARTMENT OF COMPUTER ENGINEERING
Fr. Conceicao Rodrigues College of Engineering
Bandra (W), Mumbai - 400050
University of Mumbai

2023-2024

This work is dedicated to my family.

I am very thankful for their motivation and support.

CERTIFICATE

This is to certify that the case study entitled **"Utilizing LLMs in Cybersecurity for Code Vulnerability Detection"** is a bonafide work of " Aditya Vyas (9238), Hitesh Sharma(9233), Atharva Pawar (9427) **"** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **Bachelor of Engineering** in **Computer Engineering** (Semester- Vii).

Prof. Supriya Kamoji
(Name and Sign)

Guide/ Supervisor

Dr. Sujata Deshmukh

Dr. S. S. Rathod

Approval Sheet

Mini Project Report Approval for B.E. (Semester-VII)

This mini-project report entitled Utilizing LLMs in Cybersecurity for Code Vulnerability Detection submitted by Aditya Vyas (9238)**,** Hitesh Sharma , Atharva Pawar (9427) is approved for the degree of Bachelor of Engineering in **Computer Engineering** (Semester-V).

Examiner 1._____

Examiner 2._____

Date:

Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date:

Aditya Vyas (9238)

Hitesh Sharma (9233)

Atharva Pawar (9427)

Abstract

The proliferation of software applications in our modern digital landscape has led to an unprecedented need for robust cybersecurity measures to protect against vulnerabilities and potential exploits. This case study explores the promising application of Large Language Models (LLMs) in enhancing code vulnerability detection within the realm of cybersecurity.

In this investigation, we examine the potential of state-of-the-art LLMs, such as GPT-2.5 etc , to automate the process of identifying and mitigating code vulnerabilities, which have become a significant concern for software developers and organizations. Furthermore, it sheds light on the ways LLMs can streamline and optimize the workflow of cybersecurity professionals by automating parts of the vulnerability detection process.

In this study we also tested various LLMs and fine tuning and their performances to get a deeper understanding.

Results from this case study demonstrate the potential for LLMs to significantly enhance code vulnerability detection in real-world scenarios. By harnessing the power of these models, organizations can bolster their cybersecurity posture and proactively identify and address vulnerabilities before they can be exploited by malicious actors.

.

Keywords:
LLM, Code vulnerability, cybersecurity

Acknowledgments

We have great pleasure in presenting the report on "Utilizing LLMs in Cybersecurity for Code Vulnerability Detection". I take this opportunity to express my sincere thanks towards the guide Prof.Supriya Kamoji, C.R.C.E, Bandra (W), Mumbai, for providing the technical guidelines, and the suggestions regarding the line of this work. We enjoyed discussing the work progress with him/her during our visits to the department.

We thank Dr. Sujata Deshmukh, Head of Computer Engineering department, Principal and the management of C.R.C.E., Mumbai for encouragement and providing necessary infrastructure for pursuing the project.

We also thank all non-teaching staff for their valuable support, to complete our project.

Date:

Aditya Vyas (9238)

Hitesh Sharma (9233)

Atharva Pawar (9427)

Chapter 1

Review of literature

Research paper	Date	Summary	Publisher	Link
DiverseVul: A New Vulnerable Source Code Dataset for Deep Learning Based Vulnerability Detection	2023	Detailed comparison of various AI techniques for code vulnerability detection	arxiv.org	https://arxiv.org/abs/2304.00409
From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy	2023	The paper discusses how llms(especially chat gpt and bard) can be used in the attack and defensive side of cybersecurity.	IEEE Access	https://ieeexplore.ieee.org/abstract/document/10198233
Impacts and Risk of Generative AI Technology on Cyber Defense	2023	This research paper investigates the cybersecurity risks of Generative AI (GenAI). GenAI's ability to create realistic content raises concerns for misuse like phishing and disinformation. The study analyzes GenAI's impact on cyberattacks and suggests defense strategies, addressing potential threats and offering proactive measures for effective cybersecurity.	Arxiv.org	https://arxiv.org/pdf/2306.13033.pdf

Static Code Analysis	Website	Static Code Analysis is vital in the Security Development Lifecycle's Implementation phase. It uses tools to find vulnerabilities, assists analysts, and complies with standards like UK Defense Standard 00-55 for safety-related defense software.	OWASP	https://owasp.org/www-community/controls/Static_Code_Analysis

The Use of Artificial Intelligence in Cybersecurity: A Review	Website	Artificial Intelligence (AI) plays a vital role in modern cybersecurity by swiftly analyzing vast datasets, identifying various cyber threats, and continuously improving to detect novel attack variants. AI-driven systems leverage advanced algorithms, including natural language processing, to detect malware behaviors and gather insights from diverse sources. This proactive approach enhances cybersecurity by prioritizing potential threats effectively, even in the face of emerging and industry-specific risks..	IEEE Computer Society	https://www.computer.org/publications/tech-news/trends/the-use-of-artificial-intelligence-in-cybersecurity

Static Code Analysis Tools: A Systematic Literature Review	2020	Static code analysis tools are essential for improving code quality by identifying defects and vulnerabilities. This paper systematically reviews these tools, categorizing them based on their language support and defect detection capabilities, contributing to a better understanding of their applications.	ResearchGate	https://www.researchgate.net/publication/347383785_Static_Code_Analysis_Tools_A_Systematic_Literature_Review