To what extent does targeted code preprocessing improve the generalization capability of Deepseek R1 for vulne

Assignee Research

May 29, 2026

Abstract

The rapid advancement of Large Language Models (LLMs) has opened up new opportunities for leveraging artificial intelligence in a variety of application domains, including cybersecurity. As the volume and sophistication of cyber threats continue to grow, there is an increasing need for intelligent systems that can automatically detect vulnerabilities, analyze malware, and respond to attacks. In this survey, we conduct a comprehensive review of the literature on the application of LLMs in cybersecurity~(LLM4Security). By comprehensively collecting over 40K relevant papers and systematically ana

1 Introduction

This paper examines: Large Language Models for Cyber Security: A Systematic Literature Review. Research question: To what extent does targeted code preprocessing improve the generalization capability of Deepseek R1 for vulnerability detection when evaluated on benchmark datasets containing low-resource programming languages?.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.8/10.

3 Results

4 papers retrieved. 6 claims extracted; 5 independently verified. Quality review score: 7.8/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The survey collected over 40,000 relevant papers on the application of LLMs in cybersecurity.	×	0.14
The survey systematically analyzed 185 papers from top security and software engineering venues.	✓	0.18
LLMs are being applied to cybersecurity tasks including vulnerability detection, malware analysis, and network intrusion	✓	0.28
Different LLM architectures, specifically encoder-only, encoder-decoder, and decoder-only, are being analyzed for applic	✓	0.22
Techniques for adapting LLMs to cybersecurity include advanced fine-tuning, prompt engineering, and external augmentatio	✓	0.26
The use of LLM-based autonomous agents represents an emerging trend shifting from single-task execution to orchestration	✓	0.17

References

- https://doi.org/10.48550/arxiv.2405.04760
- https://doi.org/10.48550/arxiv.2406.07467
- https://doi.org/10.3390/app15126651