DeepSeek R1 and CodeLlama False Positive Rates on Big-Vul Buffer Overflow Detection

Assignee Research

June 4, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: What is the comparative false positive rate of Deepseek R1 versus CodeLlama on buffer overflow vulnerabilities within the Big-Vul benchmark under varying context window sizes. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: LLMs in Disease Diagnosis: A Comparative Study of DeepSeek-R1 and O3 Mini Across Chronic Health Conditions. Research question: What is the comparative false positive rate of Deepseek R1 versus CodeLlama on buffer overflow vulnerabilities within the Big-Vul benchmark under varying context window sizes?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.5/10.

3 Results

14 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 2.5/10.

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.

References

- http://arxiv.org/abs/2503.02032v1
- http://arxiv.org/abs/2503.10486v2
- http://arxiv.org/abs/2012.15116v1