

Llama3 and DeepSeek R1 Inference Efficiency for Vulnerability Detection Under Adversarial Noise

Assignee Research

June 4, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: How does the inference efficiency of Llama3 compare to Deepseek R1 for vulnerability detection on the Big-Vul dataset when evaluated under varying levels of adversarial noise, measured in terms of. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: PPTC-R benchmark: Towards Evaluating the Robustness of Large Language Models for PowerPoint Task Completion. Research question: How does the inference efficiency of Llama3 compare to Deepseek R1 for vulnerability detection on the Big-Vul dataset when evaluated under varying levels of adversarial noise, measured in terms of latency and memory consumption per 1000 lines of code?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.5/10.

3 Results

16 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.5/10.

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/2509.17337v1>
- <http://arxiv.org/abs/2403.03788v1>
- <http://arxiv.org/abs/2601.08691v1>