Model Size and Adversarial Robustness in Llama3 and Codestral on Code Weakness Taxonomies

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: What is the correlation between model size and robustness against adversarial code perturbations when evaluating Llama3 and Codestral on common weakness taxonomies. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Benchmarking LLAMA Model Security Against OWASP Top 10 For LLM Applications. Research question: What is the correlation between model size and robustness against adversarial code perturbations when evaluating Llama3 and Codestral on common weakness taxonomies?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.2/10.

3 Results

15 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 5.2/10.

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/2504.16584v1
- http://arxiv.org/abs/2103.15670v3
- http://arxiv.org/abs/2601.19970v1