Scaling Model Parameters Enhances Zero-Shot Vulnerability Severity Consistency Across Languages

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: Does scaling model parameters improve the consistency of vulnerability severity scoring across different programming languages in zero-shot settings. 15 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 1.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Cross-Lingual Consistency of Factual Knowledge in Multilingual Language Models. Research question: Does scaling model parameters improve the consistency of vulnerability severity scoring across different programming languages in zero-shot settings?.

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 1.5/10.

3 Results

15 papers retrieved. 15 claims extracted; 0 independently verified. Quality review score: 1.5/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
The capital of People's Republic of China is Beijing.	×	0.02
The original language of The Godfather is Italian.	×	0.02
The BLOOM-3b model outputs consistently correct completions of the first prompt when queried in English, Spanish, and Vi	×	0.01
The BLOOM-3b model outputs consistent, though wrong, answers to the second query in English, Spanish, and Vietnamese (bu	×	0.03
The study of cross-lingual consistency (CLC) is important for at least two reasons: Firstly, true knowledge of a fact im	×	0.14
The multilingual BLOOM-3b model (Scao et al., 2022) outputs consistently correct completions of the first prompt when qu	×	0.03
The multilingual BLOOM-3b model outputs consistent, though wrong, answers to the second query in English, Spanish, and V	×	0.04
The RoBERTa model has a CLC accuracy of 32.0 for English.	×	0.05
The mT5-large model has a CLC accuracy of 32.5 for French.	×	0.05
The LLaMA-7b model has a CLC accuracy of 28.5 for English.	×	0.05
The BLOOMZ-3b model has a CLC accuracy of 24.3 for English.	×	0.05
The BLOOM-3b model has a CLC accuracy of 25.2 for English.	×	0.05
The BLOOM-1.7b model has a CLC accuracy of 24.8 for English.	×	0.05
The BLOOM-1.1b model has a CLC accuracy of 23.9 for English.	×	0.05
The BLOOM-560m model has a CLC accuracy of 23.1 for English.	×	0.05

References

- http://arxiv.org/abs/2310.10378v5

- http://arxiv.org/abs/2605.14890v2
- http://arxiv.org/abs/2312.16098v1