

Generalization Capabilities of Sub-10B and 10B+ Multilingual Code Language Models on CLCDetector

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 9 peer-reviewed papers addressing the following research question: What is the impact of model size (sub-10B vs. 10B+) on the generalization capability of multilingual code LMs in detecting CLBs, as measured by robustness scores on unseen language pairs in the. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Fine-Tuning Code Language Models to Detect Cross-Language Bugs. Research question: What is the impact of model size (sub-10B vs. 10B+) on the generalization capability of multilingual code LMs in detecting CLBs, as measured by robustness scores on unseen language pairs in the CLCDetector benchmark?.

2 Methodology

Systematic literature search across multiple databases yielded 9 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.7/10.

3 Results

9 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.7/10.

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/2507.21954v2>
- <http://arxiv.org/abs/2402.17954v3>
- <http://arxiv.org/abs/2606.05846v1>