

# Synthetic Data Augmentation Effects on Cross-Language Code Security in Sub-10B SLMs

Assignee Research

June 8, 2026

## Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How does the integration of synthetic data augmentation impact the cross-language code security performance of SLMs under 10B parameters, measured by pass@1 on benchmarks like SIMCOPILOTP. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Case Study: Fine-tuning Small Language Models for Accurate and Private CWE Detection in Python Code. Research question: How does the integration of synthetic data augmentation impact the cross-language code security performance of SLMs under 10B parameters, measured by pass@1 on benchmarks like SIMCOPILOTP?.

## 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 3.8/10.

## 3 Results

14 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.8/10.

## 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/2510.21391v1>
- <http://arxiv.org/abs/2411.15497v3>
- <http://arxiv.org/abs/2504.16584v1>