CodeT5 Performance on MBPP: Procedural vs. Natural Language Pretraining Effects

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How does the performance of CodeT5 on MBPP vary when pretrained on procedural data versus natural language, when fine-tuned on Rust and evaluated on Python, measured by execution accuracy. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.7/10. 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 performance of CodeT5 on MBPP vary when pretrained on procedural data versus natural language, when fine-tuned on Rust and evaluated on Python, measured by execution accuracy?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.7/10.

3 Results

13 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 5.7/10.

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/2402.11690v1
- http://arxiv.org/abs/2504.16584v1
- http://arxiv.org/abs/2601.21725v2