CodeT5 Pass@1 Performance on Self-Invoking vs Standard Code Generation Tasks

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: How does the pass@1 performance of CodeT5 on self-invoking code generation tasks in HumanEval Pro compare to its performance on standard HumanEval when scaling from small to large parameter counts. 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: Is Self-Repair a Silver Bullet for Code Generation?. Research question: How does the pass@1 performance of CodeT5 on self-invoking code generation tasks in HumanEval Pro compare to its performance on standard HumanEval when scaling from small to large parameter counts?.

2 Methodology

Systematic literature search across multiple databases yielded 10 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

10 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/2604.21950v1
- http://arxiv.org/abs/2412.21199v2
- http://arxiv.org/abs/2306.09896v5