

Scaling Procedural Pretraining Data and CodeT5 Robustness on MBPP Under Identifier Perturbations

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: How does increasing the volume of procedural pretraining data affect CodeT5’s exact match scores on the MBPP dataset when subjected to identifier renaming perturbations. 7 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Empirical Study of Co-Renamed Identifiers. Research question: How does increasing the volume of procedural pretraining data affect CodeT5’s exact match scores on the MBPP dataset when subjected to identifier renaming perturbations?.

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 3.2/10.

3 Results

10 papers retrieved. 7 claims extracted; 0 independently verified. Quality review score: 3.2/10.

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 class <code>MetricType</code> is renamed to <code>MetricAttribute</code> in the commit <code>3ccd7a1</code> .	×	0.05
The word 'type' in the identifiers is replaced with 'attribute' in the commit <code>3ccd7a1</code> .	×	0.05
The source code editors' Replace All feature can replace all the matched strings simultaneously.	×	0.03
The rename refactoring tools provided by IDEs such as Eclipse and IntelliJ IDEA can limit the identifiers to be renamed	×	0.08
Developers cannot change multiple identifiers to be renamed at once using rename refactoring tools provided by IDEs.	×	0.12
Identifiers account for approximately 70% of the total program content.	×	0.07
Developers can facilitate knowledge transfer among developers by naming identifiers appropriately that reflect their int	×	0.05

References

- <http://arxiv.org/abs/2601.21725v2>
- <http://arxiv.org/abs/2005.07866v1>
- <http://arxiv.org/abs/2212.02035v1>