

Contrastive vs. MLM Pretraining in CodeT5: Perturbation Success on CWE-200

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: How does the targeted perturbation success rate of contrastively pretrained CodeT5 compare to MLM-based CodeT5 on the CWE-200 benchmark. 8 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Detect-Localize-Repair: A Unified Framework for Learning to Debug with CodeT5. Research question: How does the targeted perturbation success rate of contrastively pretrained CodeT5 compare to MLM-based CodeT5 on the CWE-200 benchmark?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.3/10.

3 Results

11 papers retrieved. 8 claims extracted; 0 independently verified. Quality review score: 3.3/10.

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
CodeT5-base (220M) is used as the foundation model for the CodeT5-DLR framework.	×	0.08
The SL-Java dataset contains 13 bug patterns with a total of 75,504 instances.	×	0.02
The SL-Java dataset is split into 52,789 training instances, 7,465 validation instances, and 15,250 test instances.	×	0.03
The ML-Python dataset is split into 132,243 training instances, 22,395 validation instances, and 35,457 test instances.	×	0.03
CodeT5-DLR achieves an F1 score of 63.46 and an FPR of 31.24 on the SL-Java dataset for function-level bug detection.	×	0.06
CodeT5-DLR achieves an F1 score of 54.83 and an FPR of 43.21 on the ML-Python dataset for function-level bug detection.	×	0.06
CodeT5-DLR-new achieves an MRR@1 of 27.67, MRR@5 of 38.38, FPR@1 of 3.01, and FPR@5 of 7.23 on the SL-Java dataset for b	×	0.06
CodeT5-DLR achieves an MRR@1 of 26.98, MRR@5 of 33.75, FPR@1 of 9.84, and FPR@5 of 38.46 on the ML-Python dataset for bu	×	0.06

References

- <http://arxiv.org/abs/2109.00859v1>
- <http://arxiv.org/abs/2407.05862v1>
- <http://arxiv.org/abs/2211.14875v3>