Synonym Substitution Attacks on CodeT5 and JaCoText Code Repair Performance

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: How do synonym substitution attacks affect the code repair capabilities of CodeT5 and JaCoText on the QuixBugs benchmark in terms of correct fix rates. 7 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Are Synonym Substitution Attacks Really Synonym Substitution Attacks?. Research question: How do synonym substitution attacks affect the code repair capabilities of CodeT5 and JaCoText on the QuixBugs benchmark in terms of correct fix rates?.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.3/10.

3 Results

4 papers retrieved. 7 claims extracted; 0 independently verified. Quality review score: 4.3/10.

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
The attack success rate on the testing set composed of 7.6K samples is 57.25%.	×	0.03
Among 26600 words that are swapped by PWWS, only 5398 (20.2%) words fall in the category of matched sense substitution.	×	0.06
A majority of 20055 (75.4%) word substitutions are mismatched sense substitutions.	×	0.02
The attack success rate for PWWS is 57.25%.	×	0.00
The attack success rate for TextFooler is 81.39%.	×	0.00
PWWS uses WordNet synonym set substitution with no additional constraints.	×	0.08
TextFooler uses Counter-fitted GloVe embedding kNN substitution with k = 50 and several constraints including USE senten	×	0.05

References

- http://arxiv.org/abs/2210.02844v3
- http://arxiv.org/abs/2303.12869v1
- http://arxiv.org/abs/1710.05833v2