Semantic Consistency of CodeT5 under Adversarial Attacks with Lisp and Prolog Pretraining

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: What is the comparative impact of Lisp and Prolog pretraining corpora on the semantic consistency scores of CodeT5-generated solutions under attack scenarios in the MBPP dataset. 8 claims were extracted from source literature; 3 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.6/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Unrestricted Adversarial Attacks on ImageNet Competition. Research question: What is the comparative impact of Lisp and Prolog pretraining corpora on the semantic consistency scores of CodeT5-generated solutions under attack scenarios in the MBPP dataset?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.6/10.

3 Results

15 papers retrieved. 8 claims extracted; 3 independently verified. Quality review score: 5.6/10.

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
Unrestricted adversarial attacks involve making large and visible modifications to an image that cause model misclassifi	✓	0.22
Unrestricted adversarial attack is a popular and practical direction but has not been studied thoroughly.	✓	0.34
The competition is held on the TianChi platform as part of the AI Security Challengers Program.	✓	0.27
The final total subjective evaluation score is determined by attack success rate and image quality.	×	0.03
A key challenge in computer vision is the lack of a precise mathematical metric of human perception.	×	0.02
The competition aims to find the smallest perturbation δ such that x + δ is misclassified by the target model F under	×	0.06
Direct optimization of the problem to find the smallest perturbation is intractable partly due to the lack of informatio	×	0.04
The problem is approximately solved by discretizing the continuous space of perturbation size into a discrete space.	×	0.02

References

- http://arxiv.org/abs/1909.08072v2
- http://arxiv.org/abs/cs/0404050v1
- http://arxiv.org/abs/2110.09903v2