Decoupled Knowledge Distillation Enhances CodeT5 Robustness to Adversarial Code

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: To what extent does DKD improve CodeT5's robustness to adversarial code examples in the MBXP benchmark compared to standard knowledge distillation techniques when evaluated using adversarial accuracy. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: DD-RobustBench: An Adversarial Robustness Benchmark for Dataset Distillation. Research question: To what extent does DKD improve CodeT5's robustness to adversarial code examples in the MBXP benchmark compared to standard knowledge distillation techniques when evaluated using adversarial accuracy metrics?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.3/10.

3 Results

12 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.3/10.

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.

References

- http://arxiv.org/abs/2305.00866v2
- http://arxiv.org/abs/2403.13322v3
- http://arxiv.org/abs/2311.01473v3