

# CodeT5 Robustness to Adversarial Prompts Under Mixed Pretraining Data Ratios

Assignee Research

June 7, 2026

## Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: What is the impact of varying the ratio of code-to-natural-language pretraining data on CodeT5's robustness to adversarial code prompts in low-resource programming languages as measured by pass@1. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Code Prompting Elicits Conditional Reasoning Abilities in Text+Code LLMs. Research question: What is the impact of varying the ratio of code-to-natural-language pretraining data on CodeT5's robustness to adversarial code prompts in low-resource programming languages as measured by pass@1 scores?.

## 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 3.7/10.

## 3 Results

10 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.7/10.

## 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/2306.11066v2>
- <http://arxiv.org/abs/2412.10008v1>
- <http://arxiv.org/abs/2401.10065v3>