Consistency Regularization Effects on CodeT5 Cross-Lingual Transfer in Low-Resource Languages

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does consistency regularization impact the cross-lingual transfer accuracy of CodeT5 on the MBPP benchmark when fine-tuned on low-resource languages like Rust versus high-resource languages like. 8 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: When Does Language Transfer Help? Sequential Fine-Tuning for Cross-Lingual Euphemism Detection. Research question: How does consistency regularization impact the cross-lingual transfer accuracy of CodeT5 on the MBPP benchmark when fine-tuned on low-resource languages like Rust versus high-resource languages like Python?.

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 3.5/10.

3 Results

12 papers retrieved. 8 claims extracted; 0 independently verified. Quality review score: 3.5/10.

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 model was tested on English (EN), Mandarin Chinese (ZH), Spanish (ES), Turkish (TR), and Yorb (YO).	×	0.05
The sequential fine-tuning approach involves learning the same task first on one language (L1) and then on a second lang	×	0.13
The model's performance was compared to monolingual baselines and simultaneous fine-tuning.	×	0.11
The PETs dataset for 2025 includes 3211 examples for ZH, 3098 for EN, 2952 for ES, 2436 for TR, and 2598 for YO.	×	0.03
XLM-R achieved an F1 score of 0.821 for EN, 0.768 for ZH, 0.878 for YO, 0.809 for TR, and 0.790 for ES.	×	0.02
mBERT achieved an F1 score of 0.791 for EN, 0.712 for ZH, 0.860 for YO, 0.800 for TR, and 0.720 for ES.	×	0.02
For sequential fine-tuning, XLM-R achieved an F1 score of 0.835 for EN $\rightarrow$ TR, 0.768 for ES & ZH, 0.9 for YO $\rightarrow$ ZH, 0.830 f	×	0.08
For sequential fine-tuning, mBERT achieved an F1 score of 0.812 for EN $\rightarrow$ TR, 0.738 for ES & ZH, 0.885 for YO $\rightarrow$ ZH, 0.817	×	0.08

References

- http://arxiv.org/abs/2508.11831v1
- http://arxiv.org/abs/2310.10378v5

- http://arxiv.org/abs/2506.15415v1