Adapter-Based Fine-Tuning and Adversarial Transferability Across Languages in PAWS-X

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: What is the impact of adapter-based fine-tuning on the transferability of adversarial examples across languages in the PAWS-X benchmark for XLM-R base models. 11 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: PAWS-X: A Cross-lingual Adversarial Dataset for Paraphrase Identification. Research question: What is the impact of adapter-based fine-tuning on the transferability of adversarial examples across languages in the PAWS-X benchmark for XLM-R base models?.

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 3.7/10.

3 Results

15 papers retrieved. 11 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.

5 Extracted Claims

Claim	Verified	Confidence
The MT system works better on Indo-European languages than on CJK.	×	0.02
The CJK family is more typologically and syntactically different from English.	×	0.06
61.7% of the examples are correct in all languages.	×	0.02
32 examples failed in all languages.	×	0.03
Most of the 32 examples that failed in all languages are hard or highly ambiguous.	×	0.02
Some of the 32 examples have incorrect gold labels or were generated incorrectly in the original PAWS data.	×	0.05
The ESIM model encodes each sentence using a BiLSTM and passes the concatenation of encodings through a feed-forward lay	×	0.03
BERT achieved state-of-the-art results on eleven natural language processing tasks.	×	0.11
Multilingual BERT is a single model trained on 104 languages.	×	0.07
The Zero Shot strategy does not involve machine translation.	×	0.11
The Merged strategy trains a multilingual model on all languages, including the original English pairs and machine-trans	×	0.12

References

- http://arxiv.org/abs/2205.00267v2
- http://arxiv.org/abs/1908.11828v1
- http://arxiv.org/abs/2404.12444v1