Instruction Tuning versus English Intermediate-Task Training for Zero-Shot Cross-Lingual Transfer on Non-Latin Script Languages

Assignee Research

July 22, 2026

Abstract

Instruction tuning has emerged as a powerful technique, significantly boosting zero-shot performance on unseen tasks. While recent work has explored cross-lingual generalization by applying instruction tuning to multilingual models, previous studies have primarily focused on English, with a limited exploration of non-English tasks. For an in-depth exploration of cross-lingual generalization in instruction tuning, we perform instruction tuning individually for two distinct language meta-datasets. Subsequently, we assess the performance on unseen tasks in a language different from the one used f

1 Introduction

This paper examines: Deep Exploration of Cross-Lingual Zero-Shot Generalization in Instruction Tuning. Research question: Does instruction tuning on multimodal datasets outperform English intermediate-task training for zero-shot cross-lingual transfer on XTREME classification tasks when evaluated on non-Latin script languages?.

2 Methodology

Systematic literature search across multiple databases yielded 8 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.4/10.

3 Results

8 papers retrieved. 14 claims extracted; 10 independently verified. Quality review score: 7.4/10.

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
Instruction tuning has been shown to enhance cross-lingual zero-shot generalization in multilingual large language model	✓	0.17
Prior investigations into cross-lingual generalization through instruction tuning have predominantly focused on English	✓	0.27
Previous research lacks validation of cross-lingual generalization of instruction tuning for languages other than Englis	✓	0.27
The study defines a new cross-lingual setting where the training and inference language differs.	×	0.14
The study conducts a comparative analysis between cross-lingual instruction tuning and monolingual instruction tuning.	×	0.12
The KORANI meta-dataset comprises 51 diverse Korean benchmarks, including 34 NLU benchmarks and 17 NLG benchmarks.	✓	0.26
KORANI surpasses the quantity of non-English benchmarks explored in previous multilingual research.	✓	0.22
The mT5 model is a publicly available multilingual model pre-trained with multiple languages.	×	0.14
The study employs mT5 models with sizes ranging from 1.3B to 13B parameters for experiments.	×	0.13
The study assesses cross-lingual generalization by instruction tuning in one language and evaluating unseen tasks in the	✓	0.19
The study introduces cross-lingual templates in the training or inference phases to investigate the advantage of instruc	✓	0.22
The models mT-Ko and mT-En are trained on KORANI and P3 datasets, respectively.	✓	0.29
The models mT-Ko-CT and mT-En-CT are trained on KORANI and P3 datasets, respectively, by incorporating cross-lingual tem	✓	0.39
The models mT-Ko-CI and mT-En-CI are trained on KORANI and P3 datasets, respectively, with original templates only, and	✓	0.37

References

- http://arxiv.org/abs/2402.14778v2
- http://arxiv.org/abs/2406.08796v1
- http://arxiv.org/abs/2005.13013v2