Intermediate-Task Training Effects on Inference Efficiency in Zero-Shot Cross-Lingual Settings on XTREME-R

Assignee Research

July 20, 2026

Abstract

Intermediate-task training—fine-tuning a pretrained model on an intermediate task before fine-tuning again on the target task—often improves model performance substantially on language understanding tasks in monolingual English settings. We investigate whether English intermediate-task training is still helpful on non-English target tasks. Using nine intermediate language-understanding tasks, we evaluate intermediate-task transfer in a zero-shot cross-lingual setting on the XTREME benchmark. We see large improvements from intermediate training on the BUCC and Tatoeba sentence retrieval tas

1 Introduction

This paper examines: Intermediate-Task Training Effects on Zero-Shot Cross-Lingual Inference Efficiency in XTREME-R. Research question: How does intermediate-task training affect inference efficiency (throughput, latency) in zero-shot cross-lingual settings on the XTREME-R benchmark when comparing multilingual models to English-only models?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.8/10.

3 Results

16 papers retrieved. 6 claims extracted; 6 independently verified. Quality review score: 8.8/10.

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
Intermediate-task training often improves model performance substantially on language understanding tasks in monolingual	✓	0.44
Using nine intermediate language-understanding tasks, we evaluate intermediate-task transfer in a zero-shot cross-lingua	✓	0.52
We see large improvements from intermediate training on the BUCC and Tatoeba sentence retrieval tasks.	✓	0.30
Research goal: How does intermediate-task training with English-only vs. multilingual NLI datasets affect the inference	✓	0.62
Autonomous synthesis report generated by Assignee Research.	✓	0.25
Tribunal consensus score: 9.3/10.	✓	0.19

References

- https://doi.org/10.5281/zenodo.21317253
- https://doi.org/10.5281/zenodo.21317254
- https://doi.org/10.5281/zenodo.21330965