

Intermediate-Task Training Effects on XTREME-R Inference Throughput in Multilingual Models

Assignee Research

July 21, 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: Impact of English Intermediate-Task Training on Multilingual Model Inference Efficiency. Research question: How does intermediate-task training on English affect the inference throughput of multilingual models on XTREME-R tasks compared to direct fine-tuning?.

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 9.1/10.

3 Results

15 papers retrieved. 7 claims extracted; 7 independently verified. Quality review score: 9.1/10.

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	✓	0.45
The study investigates whether English intermediate-task training is still helpful on non-English target tasks.	✓	0.35
Nine intermediate language-understanding tasks were used to evaluate intermediate-task transfer in a zero-shot cross-lin	✓	0.43
Large improvements from intermediate training were observed on the BUCC and Tatoeba sentence retrieval tasks.	✓	0.25
The research goal is to determine if intermediate-task training on English data impacts the inference efficiency (latenc	✓	0.45
The autonomous synthesis report was generated by Assignee Research.	✓	0.25
The tribunal consensus score is 9.3/10.	✓	0.19

References

- <https://doi.org/10.5281/zenodo.21392598>
- <https://doi.org/10.5281/zenodo.21322488>
- <https://doi.org/10.5281/zenodo.21322487>