

****Intermediate Task Ordering and Zero-Shot Performance on XTREME-R****

Assignee Research

July 18, 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: XLM-R Performance Degradation in Adversarial Cross-Lingual Logical Inference: Sequential vs. Interleaved Task Order Effects. Research question: How does the choice of intermediate task ordering (sequential vs. interleaved) affect zero-shot performance on XTREME-R compared to random task ordering?.

2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.5/10.

3 Results

10 papers retrieved. 3 claims extracted; 3 independently verified. Quality review score: 8.5/10.

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
Intermediate-task training often improves model performance substantially on language understanding tasks in monolingual	✓	0.39
Using nine intermediate language-understanding tasks, intermediate-task transfer in a zero-shot cross-lingual setting on	✓	0.43
The research goal is to determine the extent to which the order of intermediate-task training (sequential vs. interleave	✓	0.55

References

- <https://doi.org/10.5281/zenodo.21049879>
- <https://doi.org/10.48550/arxiv.2303.04226>
- <https://doi.org/10.5281/zenodo.21049880>