Scaling Behavior of XGLM Zero-Shot Cross-Lingual Transfer on Indonesian XNLI

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: How does the scaling behavior of XGLM's zero-shot cross-lingual transfer accuracy on Indonesian XNLI tasks differ when fine-tuned on low-resource languages (e.g., Swahili) versus high-resource. 11 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.0/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: DeFTX: Denoised Sparse Fine-Tuning for Zero-Shot Cross-Lingual Transfer. Research question: How does the scaling behavior of XGLM's zero-shot cross-lingual transfer accuracy on Indonesian XNLI tasks differ when fine-tuned on low-resource languages (e.g., Swahili) versus high-resource languages (e.g., Spanish)?.

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 3.0/10.

3 Results

15 papers retrieved. 11 claims extracted; 0 independently verified. Quality review score: 3.0/10.

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
NusaX is an SA benchmark covering five low-resource Indonesian languages.	×	0.06
There are 400 test samples in each language in NusaX to be classified as either positive, negative or neutral.	×	0.02
AmericasNLI is an extension of XNLI to 10 low-resource indigenous languages of the Americas with 750 test samples each.	×	0.07
For XLM-RBASE, DEFT-X consistently outperforms the baselines MAD-X and LT-SFT.	×	0.05
For XLM-RBASE, the best-performing configuration of DEFT-X, with rl = rt = 100, surpasses MAD-X and LT-SFT with average	×	0.02
For XLM-RLARGE, DEFT-X outperforms the baseline LT-SFT on various settings.	×	0.04
For XLM-RBASE, with rl = 200 and rt = 90% variance, DEFT-X surpasses MAD-X and LT-SFT by 1.8% and 0.3%, respectively, on	×	0.02
DEFT-X maintains comparable performance even without source language initialization.	×	0.07
Higher-order components contain useful information, making it essential to retain them after denoising.	×	0.01
Magnitude pruning, when combined with sparse fine-tuning, effectively refines the parameter selection beyond denoising,	×	0.11
Applying magnitude pruning alone without sparse fine-tuning results in a substantial drop in performance.	×	0.13

References

- http://arxiv.org/abs/2504.09645v1
- http://arxiv.org/abs/2505.15090v1
- http://arxiv.org/abs/2301.06527v1