Zero-Shot Cross-Lingual Transfer Scaling of XGLM on Indonesian XNLI Across Model Sizes

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: How does the zero-shot cross-lingual transfer accuracy of XGLM on Indonesian XNLI tasks scale relative to English as model size increases from 564M to 7.5B parameters. 12 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Revisiting the Primacy of English in Zero-shot Cross-lingual Transfer. Research question: How does the zero-shot cross-lingual transfer accuracy of XGLM on Indonesian XNLI tasks scale relative to English as model size increases from 564M to 7.5B parameters?.

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 4.8/10.

3 Results

15 papers retrieved. 12 claims extracted; 2 independently verified. Quality review score: 4.8/10.

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
The gap between English and the other two transfer languages becomes even more salient.	×	0.11
enMT scores +2.1 F1 higher than the original dataset fiO.	×	0.03
Under-trained mT5 closes or reverses the gap.	×	0.04
When mT5 is under-trained with 32B tokens (instead of 1T as the published model), the gap between English and other sour	×	0.06
Languages with non-Latin scripts have non-negligible proportions of ASCII-only answers (with at least one letter) in the	×	0.04
Some of these ASCII-only answers are English-centric entities that cannot be translated.	×	0.05
Using a uniform sampling distribution across pre-training languages, the gap between English and other fine-tuning langu	×	0.07
English was out-performed by other source languages in 5/6 experimental settings.	×	0.05
Multilingual training data was obtained by machine-translating an originally human-curated dataset (most often in Englis	×	0.06
High-resource languages are likely to have good translation systems and therefore merely appear to outperform others on	✓	0.19
XNLI extends the English MultiNLI dataset to 15 languages, including low-resource ones such as Swahili and Urdu; trainin	✓	0.20
PAWS-X is a binary classification task for paraphrase identification. Its 6 training sets were machine-translated from t	×	0.06

References

- http://arxiv.org/abs/2106.16171v1
- http://arxiv.org/abs/2306.12916v3
- http://arxiv.org/abs/2009.05713v1