Scaling Multilingual XGLM for Educational Dialogue Act Classification in Indonesian and English

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 8 peer-reviewed papers addressing the following research question: How does the multilingual capability of XGLM-564M scale with model size (e.g., 564M vs. 1.7B) for educational dialogue act classification, measured by accuracy differences between Indonesian and. 17 claims were extracted from source literature; 4 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Improving Indonesian Text Classification Using Multilingual Language Model. Research question: How does the multilingual capability of XGLM-564M scale with model size (e.g., 564M vs. 1.7B) for educational dialogue act classification, measured by accuracy differences between Indonesian and English benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 8 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.5/10.

3 Results

8 papers retrieved. 17 claims extracted; 4 independently verified. Quality review score: 4.5/10.

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
LSTM with word embedding enhanced with paragraph vector was used by Crisdayanti and Purwarianti in their experiments on	×	0.04
Ibrohim and Budi conducted experiments on multi-label hate speech and abusive language in Indonesian Twitter using rando	×	0.08
Previous works on text classification using multilingual language models have focused on its zero-shot capability across	✓	0.18
Using its largest multilingual model, the model showed that it is very competitive and even outperforming the monolingua	×	0.11
There is no research in Indonesian text classification using multilingual language models, especially in sentiment analy	✓	0.37
The latest Indonesian text classification model has been developed using representation from word embeddings and sequent	×	0.13
The combination of word embedding with RNN, LSTM, or GRU as the model has been very successful but still possesses some	×	0.01
BERT uses Transformer architecture and is pretrained on millions of texts with masked language model (MLM) objective.	×	0.03
BERT has proven massively successful, topping various benchmarks and significantly improving the performance from previo	×	0.07
Labeled Indonesian text data is scarce in comparison to English text data.	✓	0.16
Cross-lingual representation of text has enabled models to do transfer learning across languages.	✓	0.16
There are two main methods of producing cross-lingual representation on shared vector space: alignment and joint optimiz	×	0.08
Farhan and Khodra used reviews crawled from TripAdvisor for their sentiment analysis dataset.	×	0.05
Crisdayanti and Purwarianti used text from Twitter, Zomato, TripAdvisor, Facebook, Instagram, and Qraved for their senti	×	0.04
For the English language dataset, Yelp data was used, which they have open-sourced.	×	0.09
Ibrohim and Budi crawled tweets from Twitter and annotated the text with the help of 30 diverse annotators for their hat	×	0.03
For the English language hate speech dataset, data from Jigsaw was used, which comes from various conversations over the	×	0.10

References

- http://arxiv.org/abs/2009.05713v1
- http://arxiv.org/abs/2304.07499v1
- http://arxiv.org/abs/2604.27439v1