# Performance of mBERT on XTREME with Intermediate Task Fine-Tuning in Low- vs. High-Resource Languages

Assignee Research

June 23, 2026

**Abstract**

Accuracy of English-language Question Answering (QA) systems has improved significantly in recent years with the advent of Transformer-based models (e.g., BERT). These models are pre-trained in a self-supervised fashion with a large English text corpus and further fine-tuned with a massive English QA dataset (e.g., SQuAD). However, QA datasets on such a scale are not available for most of the other languages. Multi-lingual BERT-based models (mBERT) are often used to transfer knowledge from high-resource languages to low-resource languages. Since these models are pre-trained with huge text corp

## 1 Introduction

This paper examines: MuCoT: Multilingual Contrastive Training for Question-Answering in Low-resource Languages. Research question: How does the performance of mBERT on the XTREME benchmark change when fine-tuned on intermediate tasks in low-resource languages versus high-resource languages?.

## 2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.6/10.

## 3 Results

14 papers retrieved. 17 claims extracted; 15 independently verified. Quality review score: 7.6/10.

## 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 ChAII dataset contains 1,114 records of context, question, answer, and its corresponding start position in the conte | ✓ | 0.26 |
| Hindi is represented predominantly in the ChAII dataset with nearly two-thirds of the records. | ✓ | 0.20 |
| The complete test dataset of ChAII has not been disclosed to the public as it is part of an ongoing Kaggle competition. | ✓ | 0.18 |
| The test split from the ChAII training data was created using Scikit-learn's train_test_split method with a test size of | ✓ | 0.21 |
| The validation split of 100 samples was obtained by applying the same train_test_split method over the filtered train sp | ✓ | 0.16 |
| Translations and transliterations of the ChAII training split are used as augmented samples for fine-tuning the QA model | ✓ | 0.15 |
| The Stanford Question Answering Dataset (SQuAD) contains 100K records of answerable question-answer pairs along with the | ✓ | 0.19 |
| SQuAD is used to pre-train the QA head added to the pre-trained mBERT model. | ✓ | 0.20 |
| AI4Bharat's IndicTrans2 is used for translation, achieving BLEU scores of 37.9 for Hindi to English and 28.6 for Tamil t | × | 0.15 |
| BLEU scores for translating English to Bengali, Marathi, Malayalam, and Telugu are 20.3, 16.1, 16.3, and 22.0, respectiv | ✓ | 0.32 |
| Nearly 500 of the ChAII instances could not be translated to English due to inconsistent translations of the same word i | × | 0.12 |
| Nearly another 200 instances were lost when translating from English to other Indian languages for the same reason. | ✓ | 0.21 |
| The Indic-trans transliteration module is used for transliteration, supporting many Indian language scripts including En | ✓ | 0.22 |
| Transliteration is performed from Hindi and Tamil to Bengali, Marathi, Malayalam, and Telugu. | ✓ | 0.17 |
| The mBERT-QA model is fine-tuned and evaluated for Indian languages Tamil and Hindi using the ChAII dataset. | ✓ | 0.25 |
| Fine-tuning the mBERT-QA model using only the training instances in the ChAII dataset is less effective due to the small | ✓ | 0.28 |
| Translation and transliteration to other languages are used as a data augmentation strategy to overcome the small number | ✓ | 0.20 |

## References


- http://arxiv.org/abs/2506.15415v1
- http://arxiv.org/abs/2204.05814v1
- http://arxiv.org/abs/2310.09917v3