

# Pre-trained Language Models for Arabic Question Answering on TyDi QA Benchmarks

Assignee Research

May 29, 2026

## Abstract

Reading comprehension systems for low-resource languages face significant challenges in handling unanswerable questions. These systems tend to produce unreliable responses when correct answers are absent from context. To solve this problem, we introduce NCTB-QA, a large-scale Bangla question answering dataset comprising 87,805 question-answer pairs extracted from 50 textbooks published by Bangladesh’s National Curriculum and Textbook Board. Unlike existing Bangla datasets, NCTB-QA maintains a balanced distribution of answerable (57.25%) and unanswerable (42.75%) questions. NCTB-QA also

## 1 Introduction

This paper examines: NCTB-QA: A Large-Scale Bangla Educational Question Answering Dataset and Benchmarking Performance. Research question: How do different pre-trained language model architectures (BERT, RoBERTa, GPT-3) perform on Arabic question answering benchmarks when evaluated on TyDi QA Arabic test set?.

## 2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.8/10.

## 3 Results

12 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.8/10.

## 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.

## References

- <http://arxiv.org/abs/2504.11972v2>
- <http://arxiv.org/abs/2603.05462v1>
- <http://arxiv.org/abs/2206.01550v1>