

Multilingual vs. Monolingual Language Models for Arabic Question Answering Performance

Assignee Research

May 29, 2026

Abstract

Question answering(QA) is one of the most challenging yet widely investigated problems in Natural Language Processing (NLP). Question-answering (QA) systems try to produce answers for given questions. These answers can be generated from unstructured or structured text. Hence, QA is considered an important research area that can be used in evaluating text understanding systems. A large volume of QA studies was devoted to the English language, investigating the most advanced techniques and achieving state-of-the-art results. However, research efforts in the Arabic question-answering progress at

1 Introduction

This paper examines: Pre-trained Transformer-Based Approach for Arabic Question Answering : A Comparative Study. Research question: What is the impact of fine-tuning multilingual language models on Arabic QA performance compared to monolingual models on the Arabic-SQuAD dataset?.

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 7.7/10.

3 Results

15 papers retrieved. 9 claims extracted; 9 independently verified. Quality review score: 7.7/10.

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 |
|--------------------------------------------------------------------------------------------------------------------------|----------|------------|
| Question answering (QA) is one of the most challenging yet widely investigated problems in Natural Language Processing (| ✓ | 0.34 |
| QA systems try to produce answers for given questions, which can be generated from unstructured or structured text. | ✓ | 0.29 |
| QA is considered an important research area that can be used in evaluating text understanding systems. | ✓ | 0.29 |
| A large volume of QA studies was devoted to the English language, investigating the most advanced techniques and achievi | ✓ | 0.35 |
| Research efforts in the Arabic question-answering progress at a considerably slower pace due to the scarcity of research | ✓ | 0.43 |
| Recently, many pre-trained language models provided high performance in many Arabic NLP problems. | ✓ | 0.33 |
| The study evaluates the state-of-the-art pre-trained transformers models for Arabic QA using four reading comprehension | ✓ | 0.39 |
| The study fine-tuned and compared the performance of the AraBERTv2-base model, AraBERTv0.2-large model, and AraELECTRA m | ✓ | 0.30 |
| The study provides an analysis to understand and interpret the low-performance results obtained by some models. | ✓ | 0.23 |

References

- <http://arxiv.org/abs/2505.19163v1>

- <http://arxiv.org/abs/1906.05394v1>
- <http://arxiv.org/abs/2111.05671v2>