

# Retrieval-Augmented 7B and 70B Model Performance on Multilingual Qur'anQA Benchmarks

Assignee Research

June 7, 2026

## Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does the performance of retrieval-augmented 7B and 70B models compare when evaluating multilingual capabilities on the Qur'anQA benchmark using F1 scores across different language variants of the. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.0/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Evaluating the Limits of Large Language Models in Multilingual Legal Reasoning. Research question: How does the performance of retrieval-augmented 7B and 70B models compare when evaluating multilingual capabilities on the Qur'anQA benchmark using F1 scores across different language variants of the Quran?.

## 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 3.0/10.

## 3 Results

12 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.0/10.

## 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/2509.22472v1>
- <http://arxiv.org/abs/2507.23334v2>
- <http://arxiv.org/abs/2503.16581v1>