

# Retrieval-Augmented 7B and 70B Models in Religious and Domain-Specific QA Benchmarks

Assignee Research

June 7, 2026

## Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: To what extent do retrieval-augmented 7B and 70B models generalize to other religious or domain-specific question-answering benchmarks (e.g., BibleQA) when measured by response accuracy and. 4 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.0/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Investigating Retrieval-Augmented Generation in Quranic Studies: A Study of 13 Open-Source Large Language Models. Research question: To what extent do retrieval-augmented 7B and 70B models generalize to other religious or domain-specific question-answering benchmarks (e.g., BibleQA) when measured by response accuracy and faithfulness metrics?.

## 2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.0/10.

## 3 Results

16 papers retrieved. 4 claims extracted; 0 independently verified. Quality review score: 4.0/10.

## 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.

## 5 Extracted Claims

Claim	Verified	Confidence
The system employs a Retrieval-Augmented Generation (RAG) architecture, combining retrieval-based and generative methods	×	0.09
The system executes semantic search and retrieval, response generation, and citations and contextualization tasks.	×	0.04
Context Relevance is evaluated using the precision@k metric, where k represents the number of top retrieved results considered	×	0.08
The dataset was chosen based on authenticity, descriptive richness, clarity and accessibility, and relevance.	×	0.04

## References

- <http://arxiv.org/abs/2508.11281v3>
- <http://arxiv.org/abs/2503.16581v1>
- <http://arxiv.org/abs/2307.12114v3>