FAIR-RAG and FARSIQA: Enhancing Llama-3-8B-128K Consistency in Multi-Track Music QA

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: Can retrieval-augmented generation (RAG) improve the consistency of Llama-3-8B-128K's responses in multi-track comparative music QA when evaluated using a novel semantic consistency metric across. The advent of Large Language Models (LLMs) has revolutionized Natural Language Processing, yet their application in high-stakes, specialized domains like religious question answering is hindered by challenges like hallucination and unfaithfulness to authoritative sources. This. 11 claims were extracted from source literature; 5 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: FARSIQA: Faithful and Advanced RAG System for Islamic Question Answering. Research question: Can retrieval-augmented generation (RAG) improve the consistency of Llama-3-8B-128K's responses in multi-track comparative music QA when evaluated using a novel semantic consistency metric across repeated queries?.

2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.2/10.

3 Results

10 papers retrieved. 11 claims extracted; 5 independently verified. Quality review score: 6.2/10.

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
FARSIQA achieves a 'Answer Correctness' score of 74.3% (as evaluated via LLM-as-Judge on multi-hop queries).	×	0.14
FARSIQA substantially outperforms standard baselines across metrics of relevance, correctness, and robustness.	×	0.04
FARSIQA is built upon the FAIR-RAG architecture, which employs a dynamic, self-correcting process.	✓	0.23
FAIR-RAG adaptively decomposes complex queries, critically assesses the sufficiency of retrieved evidence, and enters an	✓	0.23
FARSIQA operates on a curated knowledge base of over one million documents from authoritative Islamic sources.	✓	0.15
FARSIQA achieves a remarkable 97.0% in Negative Rejection, a 40-point improvement over standard baselines.	✓	0.21
FARSIQA achieves a high Answer Correctness score of 74.3%.	✓	0.16
The knowledge base includes approximately 431,000 unique documents from eleven reputable online Persian Islamic encyclop	×	0.07
The knowledge base includes approximately 304,000 question-answer pairs from authoritative Q&A websites.	×	0.05
All sources were crawled ethically, respecting copyrights and ensuring diversity across Islamic perspectives to mitigate	×	0.02
The preprocessing pipeline includes extracting main textual content and applying a recursive chunking strategy for encyc	×	0.02

References

- http://arxiv.org/abs/2510.25621v1
- http://arxiv.org/abs/2507.23334v2
- http://arxiv.org/abs/2508.05197v2