SOVEREIGN: To what extent does the choice of retriever (BM25 vs. dense passage retriever vs. LLM-based re-ranker) impact

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 28, 2026

Abstract

Large Language Models (LLMs) showcase impressive capabilities but encounter challenges like hallucination, outdated knowledge, and non-transparent, untraceable reasoning processes. Retrieval-Augmented Generation (RAG) has emerged as a promising solution by incorporating knowledge from external databases. This enhances the accuracy and credibility of the generation, particularly for knowledge-intensive tasks, and allows for continuous knowledge updates and integration of domain-specific information. RAG synergistically merges LLMs' intrinsic knowledge with the vast, dynamic repositories of exte

1 Introduction

Analysis of: Retrieval-Augmented Generation for Large Language Models: A Survey. Research goal: To what extent does the choice of retriever (BM25 vs. dense passage retriever vs. LLM-based re-ranker) impact multi-hop reasoning accuracy in RAG systems on HotpotQA and MuSiQue under varying passage count constraints?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score $<$ 6.5.

3 Results

10 papers retrieved. 5 claims extracted, 5 verified. Tribunal: 8.2/10 → APPROVE (revision_round=0). Policy: AUTO_APPROVE.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

5 Extracted Claims

Claim	Verified	Confidence
Retrieval-Augmented Generation (RAG) has emerged as a promising solution by incorporating knowledge from external databa	✓	0.32
RAG enhances the accuracy and credibility of generation, particularly for knowledge-intensive tasks	✓	0.23
RAG allows for continuous knowledge updates and integration of domain-specific information	✓	0.23
RAG synergistically merges LLMs' intrinsic knowledge with the vast, dynamic repositories of external databases	✓	0.31
The paper introduces up-to-date evaluation framework and benchmark	✓	0.20

References

- https://doi.org/10.1162/tacl_a_00021
- https://doi.org/10.3390/electronics14112102
- https://doi.org/10.48550/arxiv.2312.10997