SOVEREIGN: How does the domain shift between synthetic training data and target benchmarks (e.g., HotPotQA vs MuSiQue) af

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Large Language Models (LLMs) showcase impressive capabilities but encounter challenges like hallucination, outdated knowledge, and non-transparent, untraceable reasoning processes.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: How does the domain shift between synthetic training data and target benchmarks (e.g., HotPotQA vs MuSiQue) affect the MRR@10 improvement from adversarial retriever training across different hop complexities?.

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

9 papers retrieved. 5 claims extracted, 4 verified. Tribunal: 7.5/10 → 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) enhances the accuracy and credibility of generation for knowledge-intensive tasks	✓	0.28
RAG synergistically merges LLMs' intrinsic knowledge with external databases	✓	0.29
This paper provides a detailed examination of RAG paradigms including Naive RAG, Advanced RAG, and Modular RAG	✓	0.22
RAG allows for continuous knowledge updates and integration of domain-specific information	✓	0.25
This paper introduces up-to-date evaluation frameworks and benchmarks	×	0.15

References

- `https://doi.org/10.3390/fi15080260`
- `https://doi.org/10.48550/arxiv.2308.07107`
- `https://doi.org/10.48550/arxiv.2312.10997`