

# SOVEREIGN: How does the number of retrieved passages per hop in iterative retrieval affect F1 and EM scores on HotPotQA w

SOVEREIGN Research Kernel

Autonomous draft — Owner review required before publication

May 29, 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Large Language Models (LLMs) showcase impressive capabilities but encounter challenges like hallucination, outdated knowledge, and non-transparent, untraceable reasoning processes.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 number of retrieved passages per hop in iterative retrieval affect F1 and EM scores on HotPotQA when using Llama-3-8B as the backbone compared to single-pass retrieval?.

## 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: 8.3/10 \$\rightarrow\$ 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	✓	0.23
Retrieval-Augmented Generation (RAG) has emerged as a promising solution by incorporating knowledge from external databases	✓	0.30
RAG synergistically merges LLMs' intrinsic knowledge with external databases	✓	0.27
RAG paradigms include Naive RAG, Advanced RAG, and Modular RAG	✓	0.18
RAG frameworks have a tripartite foundation including retrieval, generation, and augmentation techniques	✓	0.18

## References

- <https://doi.org/10.48550/arxiv.2402.06196>
- <https://doi.org/10.48550/arxiv.2405.14831>
- <https://doi.org/10.48550/arxiv.2312.10997>