

SOVEREIGN: How does the diversity-weight parameter in Vendi-RAG influence retrieval latency and EM score when evaluated o

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Large Language Models (LLMs) showcase impressive capabilities but encounter challenges like hallucination, outdated knowledge, and non-transparent, untraceable reasoning processes.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 diversity-weight parameter in Vendi-RAG influence retrieval latency and EM score when evaluated on the HotpotQA dataset with FLAN-T5-xl as the generator?.

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. 7 claims extracted, 7 verified. Tribunal: 9.2/10 \$\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 language model generation, particularly fo	✓	0.28
RAG allows for continuous knowledge updates and integration of domain-specific information	✓	0.25
RAG synergistically merges LLMs' intrinsic knowledge with vast, dynamic repositories of external databases	✓	0.33
Large Language Models (LLMs) encounter challenges like hallucination, outdated knowledge, and non-transparent, untraceab	✓	0.33
RAG has emerged as a promising solution by incorporating knowledge from external databases	✓	0.28
This paper offers a detailed examination of the progression of RAG paradigms, encompassing the Naive RAG, the Advanced R	✓	0.33
This paper meticulously scrutinizes the tripartite foundation of RAG frameworks, which includes the retrieval, the gener	✓	0.31

References

- <https://doi.org/10.48550/arxiv.2110.08207>
- <https://doi.org/10.1109/access.2021.3140175>
- <https://doi.org/10.48550/arxiv.2312.10997>