Semantic Diversity in Retrieval-Augmented Generation for Multi-Hop Question Answering

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: Does increasing the semantic diversity of retrievers in a RAG system improve TriviaQA answer accuracy compared to homogeneous retriever setups. 12 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.4/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Vendi-RAG: Adaptively Trading-Off Diversity And Quality Significantly Improves Retrieval Augmented Generation With LLMs. Research question: Does increasing the semantic diversity of retrievers in a RAG system improve TriviaQA answer accuracy compared to homogeneous retriever setups?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.4/10.

3 Results

15 papers retrieved. 12 claims extracted; 2 independently verified. Quality review score: 4.4/10.

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
Experiments were conducted on three multi-hop QA benchmark datasets: MuSiQue, HotpotQA, and 2WikiMultiHopQA.	✓	0.16
The sensitivity analysis of the VSR process used 100 randomly sampled queries from the dataset.	×	0.04
Setting the parameter s = 0.0 represents a baseline pure similarity search scenario relying exclusively on cosine simila	×	0.04
In the sensitivity analysis, increasing the parameter s from 0.0 to 1.0 causes both Kendall's τ and Spearman's ρ to decr	×	0.02
At parameter s = 0.2, the Kendall's τ value is 0.797 and the Spearman's ρ value is 0.828.	×	0.01
At parameter s = 1.0, the Kendall's τ value is 0.074 and the Spearman's ρ value is 0.078.	×	0.01
Higher s values in the Vendi-RAG retrieval process promote retrieval diversity by prioritizing documents that may be les	×	0.13
The Vendi Score (VS) explicitly quantifies semantic diversity in a set of documents.	×	0.12
The Vendi Score attains its maximum value n when all documents in the set are orthogonal (fully diverse).	×	0.05
Standard similarity search (SS) often results in redundant documents with high similarity.	×	0.05
Maximal Marginal Relevance (MMR) struggles to capture global semantic diversity compared to the Vendi Score approach.	×	0.09
Vendi-RAG variants (s=0.8, s=1.0, s=0.3) and Adaptive-RAG were evaluated on the 2WikiMultiHopQA, HotpotQA, and MuSiQue d	✓	0.17

References

- http://arxiv.org/abs/2404.07220v2
- http://arxiv.org/abs/2510.25621v1
- http://arxiv.org/abs/2502.11228v2