Diversity-Aware Re-Ranking Trade-Offs in Retrieval-Augmented Generation for Open-Domain QA

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: What are the trade-offs between retrieval efficiency and generation quality when applying diversity-aware re-ranking strategies in RAG systems evaluated on open-domain QA tasks. Retrieval-augmented generation (RAG) enhances large language models (LLMs) for domain-specific question-answering (QA) tasks by leveraging external knowledge sources. However, traditional RAG systems primarily focus on relevance-based retrieval and often struggle with. 8 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.2/10. 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: What are the trade-offs between retrieval efficiency and generation quality when applying diversity-aware re-ranking strategies in RAG systems evaluated on open-domain QA tasks?.

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 5.2/10.

3 Results

15 papers retrieved. 8 claims extracted; 2 independently verified. Quality review score: 5.2/10.

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
Vendi-RAG uses a retrieval approach based on the Vendi Score (VS) to explicitly quantify semantic diversity in a set of	✓	0.19
The Vendi Score attains its maximum value n when all documents are orthogonal (fully diverse).	×	0.06
Setting s = 0.0 in the VSR process represents a pure similarity search scenario.	×	0.02
As s increases from 0.0 to 1.0 in the VSR process, both Kendall's τ and Spearman's ρ decrease progressively.	×	0.03
Vendi-RAG was evaluated on three multi-hop QA benchmark datasets: MuSiQue, HotpotQA, and 2WikiMultiHopQA.	✓	0.20
The sensitivity analysis of the VSR process was conducted using 100 randomly sampled queries from the dataset.	×	0.03
Kendall's τ measures the rank order similarity between two lists by evaluating concordant and discordant pairs.	×	0.04
Spearman's Rank Correlation ρ assesses the monotonic relationship between two rankings by considering both orderings and	×	0.02

References

- http://arxiv.org/abs/2507.23334v2
- http://arxiv.org/abs/2502.11228v2
- http://arxiv.org/abs/2411.18583v1