Vendi-RAG Diversity-Weight Impact on ELI5 Performance with Sparse and Dense Retrievers

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 1 peer-reviewed paper addressing the following research question: How does the diversity-weight parameter in Vendi-RAG affect its performance on the ELI5 dataset when using a sparse retriever versus a dense retriever, measured by ROUGE-L scores. Questa tesi affronta il problema della ridondanza informativa nei sistemi di Retrieval-Augmented Generation (RAG), dove i modelli di linguaggio di grandi dimensioni vengono supportati da documenti recuperati da collezioni esterne. La presenza di passaggi duplicati o parafrasati. 8 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Beyond Redundancy: Embedding-Aware Novelty Reranking in Retrieval-Augmented Generation. Research question: How does the diversity-weight parameter in Vendi-RAG affect its performance on the ELI5 dataset when using a sparse retriever versus a dense retriever, measured by ROUGE-L scores?.

2 Methodology

Systematic literature search across multiple databases yielded 1 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.7/10.

3 Results

1 papers retrieved. 8 claims extracted; 1 independently verified. Quality review score: 5.7/10.

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
The thesis addresses the problem of informational redundancy in Retrieval-Augmented Generation (RAG) systems.	✓	0.22
Redundant or paraphrased passages limit factual coverage and reduce the effectiveness of generated responses in RAG syst	×	0.03
The approach studied involves novelty-based semantic reranking using clustering in dense embedding spaces.	×	0.05
The goal is to promote content diversity presented to the model, reducing redundancy and favoring the inclusion of compl	×	0.01
Different supervision strategies (lexical and semantic) are evaluated, including fine-tuning the Set-Encoder model with	×	0.05
The evaluation is conducted on standard benchmarks (MMLU, GPQA) and the TREC RAG Track 2024, using accuracy and nugget-b	×	0.10
The results show that the benefits of novelty-aware reranking are particularly evident in evaluation contexts that prior	×	0.05
The thesis confirms the potential of embedding-aware approaches in building more informative and effective RAG pipelines	×	0.06

References

- https://openalex.org/W7113598051