

# Vendi-RAG Diversity-Weight Impact on FLAN-T5-xl Alignment with Human Preferences

Assignee Research

June 1, 2026

## Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How does the diversity-weight parameter in Vendi-RAG influence the alignment of FLAN-T5-xl outputs with human preferences on the TruthfulQA benchmark compared to BM25 retrieval. 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. 13 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. 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: How does the diversity-weight parameter in Vendi-RAG influence the alignment of FLAN-T5-xl outputs with human preferences on the TruthfulQA benchmark compared to BM25 retrieval?.

## 2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.5/10.

### **3 Results**

13 papers retrieved. 13 claims extracted; 2 independently verified. Quality review score: 4.5/10.

### **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 was evaluated on three multi-hop QA benchmark datasets: MuSiQue, HotpotQA, and 2WikiMultiHopQA.	✓	0.21
The sensitivity analysis of the VSR process was conducted using 100 randomly sampled queries from the dataset.	×	0.03
Setting $s = 0.0$ serves as a baseline representing a pure similarity search scenario.	×	0.03
Kendall’s $\tau$ and Spearman’s $\rho$ were used to quantify deviations from the baseline in the sensitivity analysis.	×	0.03
As $s$ increases from 0.0 to 1.0, both Kendall’s $\tau$ and Spearman’s $\rho$ decrease progressively.	×	0.03
Vendi-RAG-4o achieved an F1-score of 69.9% on the 2WikiMultiHopQA dataset.	×	0.07
Adaptive-RAG-4o achieved an F1-score of 60.1% on the 2WikiMultiHopQA dataset.	×	0.06
Vendi-RAG-4o achieved an Exact Match score of 63.4% on the 2WikiMultiHopQA dataset.	×	0.07
Adaptive-RAG-4o achieved an Exact Match score of 56.5% on the 2WikiMultiHopQA dataset.	×	0.05
Vendi-RAG-4o achieved an Accuracy of 63.4% on the 2WikiMultiHopQA dataset.	×	0.09
Adaptive-RAG-4o achieved an Accuracy of 59.3% on the 2WikiMultiHopQA dataset.	×	0.07
Vendi-RAG uses the Vendi Score (VS) to quantify semantic diversity in a set of documents.	✓	0.17
The Vendi Score (VS) attains its maximum value $n$ when all documents are orthogonal (fully diverse).	×	0.06

## References

- <http://arxiv.org/abs/2503.16581v1>
- <http://arxiv.org/abs/2502.11228v2>
- <http://arxiv.org/abs/2504.01346v4>