Vendi-RAG Diversity-Weight Parameter Effects on ELI5 Factuality and Coherence

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: How does the diversity-weight parameter in Vendi-RAG influence the model's performance on the ELI5 dataset when evaluated using human judgments for factuality and coherence, compared to automated. While humans increasingly rely on large language models (LLMs), they are susceptible to generating inaccurate or false information, also known as "hallucinations". Technical advancements have been made in algorithms that detect hallucinated content by assessing the factuality of. 11 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Facilitating Human-LLM Collaboration through Factuality Scores and Source Attributions. Research question: How does the diversity-weight parameter in Vendi-RAG influence the model's performance on the ELI5 dataset when evaluated using human judgments for factuality and coherence, compared to automated metrics like ROUGE-L?.

2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.2/10.

3 Results

10 papers retrieved. 11 claims extracted; 0 independently verified. Quality review score: 4.2/10.

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
The analyses were conducted using generalized linear mixed-effects models, with one model for each dependent variable: p	×	0.14
In both models, the categorical independent variable Design Strategy was treatment-coded with the Baseline design set as	×	0.05
Participant ID was a random variable.	×	0.00
Following the omnibus models, pairwise contrasts were conducted to explore comparisons between each pair of Design Strat	×	0.02
The goal of the current study was to understand how to present information about factuality and source attributions to a	×	0.11
For the controlled study, we selected six different designs for representing a factuality score and two different design	×	0.06
This selection was based on a pilot study in which we interviewed ten participants about their preferences on numerous d	×	0.06
There were 104 participants in the controlled study, who were employees of a large, multinational technology company.	×	0.04
Participant recruitment was advertised widely within the company on 25 internal Slack channels spanning multiple divisio	×	0.01
Participants' work locations consisted of 20 unique countries, and job roles spanned a wide array of disciplines, includ	×	0.02
Participants had a range of experiences with AI as a technology, with some having heard of it and others having extensiv	×	0.04

References

- http://arxiv.org/abs/2405.20434v1
- http://arxiv.org/abs/2502.11228v2
- http://arxiv.org/abs/2506.16345v1