

SOVEREIGN: How does Vendi-RAG’s iterative retrieval process compare to fixed-top-k retrieval on the 2WikiMultihop benchma

SOVEREIGN Research Kernel

Autonomous draft — Owner review required before publication

May 29, 2026

Abstract

In the last few years, the deep learning (DL) computing paradigm has been deemed the Gold Standard in the machine learning (ML) community. Moreover, it has gradually become the most widely used computational approach in the field of ML, thus achieving outstanding results on several complex cognitive tasks, matching or even beating those provided by human performance. One of the benefits of DL is the ability to learn massive amounts of data. The DL field has grown fast in the last few years and it has been extensively used to successfully address a wide range of traditional applications. More i

1 Introduction

Analysis of: Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Research goal: How does Vendi-RAG’s iterative retrieval process compare to fixed-top-k retrieval on the 2WikiMultihop benchmark when evaluated using F1 and exact match scores under varying levels of injected irrelevant context?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score < 6.5.

3 Results

6 papers retrieved. 0 claims extracted, 0 verified. Tribunal: 5.3/10 → RE-
VISE (revision_round=1). Policy: ESCALATE_TO_OWNER.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv
Relevance ranking is query-dependent. Tribunal consensus is LLM-based
and prompt-sensitive.

References

- <https://doi.org/10.1186/s40537-021-00444-8>
- <https://doi.org/10.1186/s40537-021-00492-0>
- <https://doi.org/10.48550/arxiv.2312.10997>