

SOVEREIGN: How does adversarial passage perturbation severity (measured via semantic similarity degradation) affect the t

SOVEREIGN Research Kernel

Autonomous draft — Owner review required before publication

May 28, 2026

Abstract

While Retrieval-Augmented Generation (RAG) mitigates hallucination and knowledge staleness in Large Language Models (LLMs), existing frameworks often falter on complex, multi-hop queries that require synthesizing information from disparate sources. Current advanced RAG methods, employing iterative or adaptive strategies, lack a robust mechanism to systematically identify and fill evidence gaps, often propagating noise or failing to gather a comprehensive context. We introduce FAIR-RAG, a novel agentic framework that transforms the standard RAG pipeline into a dynamic, evidence-driven reasoning

1 Introduction

Analysis of: FAIR-RAG: Faithful Adaptive Iterative Refinement for Retrieval-Augmented Generation. Research goal: How does adversarial passage perturbation severity (measured via semantic similarity degradation) affect the token-level latency and end-to-end throughput of RAG pipelines using 7B, 13B, and 70B LLMs on the HotpotQA multi-hop benchmark?.

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

11 papers retrieved. 8 claims extracted, 1 verified. Tribunal: 6.2/10 → REVISE (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.

5 Extracted Claims

Claim	Verified	Confidence
FAIR-RAG uses Reciprocal Rank Fusion (RRF) to re-rank documents and combine rankings from multiple retrieval methods.	×	0.05
The RRF algorithm produces a single, robustly ranked list of the top-5 most relevant documents as candidate evidence.	×	0.03
In FAIR-RAG, candidate evidence is filtered by an LLM agent that evaluates each document's utility with respect to the o	×	0.06
Documents deemed irrelevant, off-topic, or only tangentially related are discarded during the evidence filtering step.	×	0.03
The Structured Evidence Assessment (SEA) module in FAIR-RAG performs granular gap analysis to generate actionable signal	✓	0.15
FAIR-RAG's SEA module uses a checklist-based methodology to deconstruct user queries into discrete informational compone	×	0.08
The FAIR-RAG framework includes an Adaptive Routing module that analyzes query complexity to determine an optimal execut	×	0.07
FAIR-RAG uses a cyclical process designed to progressively build and validate context for complex queries.	×	0.07

References

- <http://arxiv.org/abs/2510.22344v1>

- <http://arxiv.org/abs/2502.11228v2>
- <http://arxiv.org/abs/2401.15391v1>