

SOVEREIGN: How does the accuracy of LLM-based multi-hop RAG systems degrade under adversarial query perturbations (e.g.,

SOVEREIGN Research Kernel

Autonomous draft — Owner review required before publication

May 28, 2026

Abstract

Multi-hop question answering is a knowledge-intensive complex problem. Large Language Models (LLMs) use their Chain of Thoughts (CoT) capability to reason complex problems step by step, and retrieval-augmentation can effectively alleviate factual errors caused by outdated and unknown knowledge in LLMs. Recent works have introduced retrieval-augmentation in the CoT reasoning to solve multi-hop question answering. However, these chain methods have the following problems: 1) Retrieved irrelevant paragraphs may mislead the reasoning; 2) An error in the chain structure may lead to a cascade of erro

1 Introduction

Analysis of: Tree of Reviews: A Tree-based Dynamic Iterative Retrieval Framework for Multi-hop Question Answering. Research goal: How does the accuracy of LLM-based multi-hop RAG systems degrade under adversarial query perturbations (e.g., synonym substitution, negation) compared to single-hop queries when evaluated on the BEIR benchmark using dense retrievers vs. sparse retrievers?.

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

3 papers retrieved. 4 claims extracted, 1 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.

5 Extracted Claims

Claim	Verified	Confidence
The Tree of Reviews (TOR) framework achieves state-of-the-art performance in both retrieval and response generation on m	✓	0.32
The TOR framework integrates a tree structure into the iterative retrieval process to mitigate the negative impact of ch	×	0.12
Tree of Reviews (TOR) framework includes two search optimization strategies: pruning and effective expansion.	×	0.06
Tree of Reviews framework uses a tree-like structure to dynamically initiate requests based on external knowledge and in	×	0.13

References

- <http://arxiv.org/abs/2404.14464v1>
- <http://arxiv.org/abs/2507.23334v2>
- <http://arxiv.org/abs/1912.02145v1>