

SOVEREIGN: Tree of Reviews: A Tree-based Dynamic Iterative Retrieval Framework for Multi-hop

SOVEREIGN Research Kernel

Autonomous draft — Owner review required before publication

May 27, 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to reason complex problems step by step, and retrieval-augmentation can effectively alleviate factual errors caused by outdated and unknown knowledge in LLMs.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 errors.

1 Introduction

Analysis of: Tree of Reviews: A Tree-based Dynamic Iterative Retrieval Framework for Multi-hop Question Answering. Research goal: What is the marginal accuracy improvement of extending context windows from 32K to 128K tokens for multi-hop reasoning in RAG systems on HotPotQA, and does this gain saturate after a specific retrieval step count (e.g., 3 steps) when comparing 7B and 70B parameter LLMs?.

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. 0 claims extracted, 0 verified. Tribunal: 3.0/10 → REJECT (revision_round=0). 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

- <http://arxiv.org/abs/2504.13684v1>
- <http://arxiv.org/abs/2510.22344v1>
- <http://arxiv.org/abs/2404.14464v1>