

SOVEREIGN: What is the impact of varying the number of retrieved passages per hop on the accuracy of multi-hop reasoning

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to reason complex problems step by step, and retrieval-augmentation can effectively alleviate factual errors caused by outdated and unknown knowledge in LLMs.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: What is the impact of varying the number of retrieved passages per hop on the accuracy of multi-hop reasoning in RAG systems evaluated on HotPotQA, and how does this compare across LLMs of different sizes (e.g., 7B vs 70B)?.

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. 4 claims extracted, 2 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
The TREE OF REVIEWS (TOR) framework achieves state-of-the-art performance in both retrieval and response generation on t	✓	0.35
The TREE OF REVIEWS (TOR) framework uses a tree-based dynamic retrieval approach where the root node is the question and	✓	0.28
The TREE OF REVIEWS (TOR) framework mitigates the negative impact associated with the inherent vulnerabilities of chain-	×	0.10
Two tree-based search optimization strategies, pruning and effective expansion, are proposed in the TREE OF REVIEWS fram	×	0.08

References

- <http://arxiv.org/abs/2404.14464v1>
- <http://arxiv.org/abs/2403.03788v1>
- <http://arxiv.org/abs/2407.08029v1>