

# Tree of Reviews vs. Chain-Based Retrieval Efficiency in Llama-3-8B at 128K Context

Assignee Research

May 30, 2026

## Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does the Tree of Reviews retrieval framework compare to chain-based retrieval in terms of computational efficiency and latency when applied to Llama-3-8B models on the MuSiQue benchmark at 128K. 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to reason complex problems step by step, and retrieval-augmentation can effectively alleviate factual errors caused by outdated and. 5 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Tree of Reviews: A Tree-based Dynamic Iterative Retrieval Framework for Multi-hop Question Answering. Research question: How does the Tree of Reviews retrieval framework compare to chain-based retrieval in terms of computational efficiency and latency when applied to Llama-3-8B models on the MuSiQue benchmark at 128K context length?.

## 2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.5/10.

### 3 Results

12 papers retrieved. 5 claims extracted; 1 independently verified. Quality review score: 4.5/10.

### 4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

### 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 proposed TREE OF REVIEWS (TOR) framework integrates a tree structure into the iterative retrieval process.	×	0.14
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
The proposed search optimization strategies demonstrate significant improvements.	×	0.04

### References

- <http://arxiv.org/abs/2404.14464v1>
- <http://arxiv.org/abs/2605.02623v1>
- <http://arxiv.org/abs/2602.00899v1>