What is the impact of context length on F1 score degradation for Llama-3-8B-128K on the MuSiQue benchmark using Tree of Reviews vs chain-based retrieval

Assignee Research

May 29, 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to reason complex problems step by step, and retrieval-augmentation can effectively alleviate factual errors caused by outdated and unknown knowledge in LLMs.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

1 Introduction

This paper examines: Tree of Reviews: A Tree-based Dynamic Iterative Retrieval Framework for Multi-hop Question Answering. Research question: What is the impact of context length on F1 score degradation for Llama-3-8B-128K on the MuSiQue benchmark using Tree of Reviews vs chain-based retrieval?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.5/10.

3 Results

13 papers retrieved. 8 claims extracted; 3 independently verified. Quality review score: 5.5/10.

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
TOR achieves state-of-the-art performance in both retrieval and response generation on three different multi-hop questio	✓	0.32
Tree of Thought (ToT) enhances the problem-solving capabilities of Large Language Models (LLMs) by introducing a tree-li	×	0.08
The tree is an efficient structure for solving complex reasoning problems.	×	0.09
TOR is the first retrieval framework that uses a tree-like structure to dynamically initiate requests based on external	×	0.14
TOR introduces a tree structure to handle each retrieved paragraph separately, alleviating the misleading effect of irre	✓	0.31
The diversity of reasoning path extension in TOR reduces the impact of a single reasoning error on the whole.	✓	0.24
TOR proposes two tree-based search optimization strategies: pruning and effective expansion.	×	0.06
Pruning and effective expansion strategies in TOR demonstrate significant improvements in reducing time overhead and inc	×	0.04

References

- http://arxiv.org/abs/2507.23334v2

- http://arxiv.org/abs/2510.14278v1
- http://arxiv.org/abs/2404.14464v1