Language Models and Multi-Hop Reasoning in Scientific Question Answering

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How do language models handle multi-hop reasoning chains in scientific question answering v9. 12 claims were extracted from source literature; 3 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.2/10. 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 do language models handle multi-hop reasoning chains in scientific question answering v9.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.2/10.

3 Results

14 papers retrieved. 12 claims extracted; 3 independently verified. Quality review score: 5.2/10.

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	×	0.07
The tree is an efficient structure for solving complex reasoning problems.	×	0.08
Asai et al. (2020) trained a retriever that dynamically retrieves information from Wikipedia graphs.	×	0.06
Some researchers decompose complex problems into a static problem tree with several subproblems.	×	0.08
The decomposition of the question and the construction of the tree lack the assistance of external knowledge and informa	×	0.06
TOR is the first to propose a retrieval framework that uses a tree-like structure to dynamically initiate requests based	×	0.14
LLMs can decide dynamically whether to initiate further retrieval and what requests to generate based on external knowle	×	0.09
TOR introduces a tree structure to handle each retrieved paragraph separately, alleviating the misleading effect of irre	✓	0.32
The diversity of reasoning path extension 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 demonstrate significant improvements in reducing time overhead and increasing	×	0.03

References

- http://arxiv.org/abs/2603.14257v1
- http://arxiv.org/abs/2510.25621v1
- http://arxiv.org/abs/2404.14464v1