Cross-Modal Retrieval Integration in RAG Systems and Multi-Hop Reasoning Accuracy

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 8 peer-reviewed papers addressing the following research question: How does the integration of cross-modal retrieval (e.g., text-image) in RAG systems affect the multi-hop reasoning accuracy of LLMs compared to purely semantic or hybrid query-based retrieval on. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Blended RAG: Improving RAG (Retriever-Augmented Generation) Accuracy with Semantic Search and Hybrid Query-Based Retrievers. Research question: How does the integration of cross-modal retrieval (e.g., text-image) in RAG systems affect the multi-hop reasoning accuracy of LLMs compared to purely semantic or hybrid query-based retrieval on benchmarks like MultiHopQA or MedMCQA?.

2 Methodology

Systematic literature search across multiple databases yielded 8 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.8/10.

3 Results

8 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.8/10.

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.

References

- http://arxiv.org/abs/2404.07220v2
- http://arxiv.org/abs/2404.14464v1
- http://arxiv.org/abs/2604.18234v1