

Retriever-Generator Co-Training for Hallucination Reduction in Knowledge-Intensive Tasks

Assignee Research

May 29, 2026

Abstract

Large Language Models (LLMs) excel in language comprehension and generation but are prone to hallucinations, producing factually incorrect or unsupported outputs. Retrieval Augmented Generation (RAG) systems address this issue by grounding LLM responses with external knowledge. This study evaluates the relationship between retriever effectiveness and hallucination reduction in LLMs using three retrieval approaches: sparse retrieval based on BM25 keyword search, dense retrieval using semantic search with Sentence Transformers, and a proposed hybrid retrieval module. The hybrid module

1 Introduction

This paper examines: Hybrid Retrieval for Hallucination Mitigation in Large Language Models: A Comparative Analysis. Research question: How does retriever-generator co-training influence factuality scores and hallucination rates in knowledge-intensive downstream applications?.

2 Methodology

Systematic literature search across multiple databases yielded 16 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

16 papers retrieved. 11 claims extracted; 2 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 hybrid retrieval system assigns a weight of 0.7 to Sparse and 0.3 to Dense for specific queries.	×	0.09
The hybrid retrieval system assigns a weight of 0.7 to Dense and 0.3 to Sparse for generic queries.	×	0.07
The hybrid retrieval system uses WordNet for query expansion.	×	0.09
The hybrid retrieval system uses Reciprocal Rank Fusion for combining results from Sparse and Dense retrievers.	✓	0.20
The hybrid retrieval system uses top-k results from Sparse and Dense retrievers.	✓	0.19
The hybrid retrieval system uses dynamic weighting for combining results from Sparse and Dense retrievers.	×	0.14
The hybrid retrieval system uses a query expansion module to address lexical chasm.	×	0.09
The retrieval phase of a LLM-driven RAG system contains two main components: the indexed database and the retriever.	×	0.05
The indexed database DB is an external knowledge-base which is a structured collection of documents $d_i \in D$, for $i = \{1, ..$	×	0.04
The retriever encodes q and all D in a vector space.	×	0.03
The retriever applies a chosen similarity function f_{sim} which computes a similarity score between two given vector repre	×	0.02

References

- <http://arxiv.org/abs/2112.08688v2>
- <http://arxiv.org/abs/2504.05324v1>
- <http://arxiv.org/abs/2503.16581v1>