

Hybrid Retrieval Effects on Factual Consistency in Table-Heavy vs. Text-Heavy Telco-DPR Subsets

Assignee Research

June 2, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How does hybrid retrieval impact factual consistency scores on table-heavy subsets of the Telco-DPR benchmark compared to text-heavy subsets. 10 claims were extracted from source literature; 4 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.9/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Telco-DPR: A Hybrid Dataset for Evaluating Retrieval Models of 3GPP Technical Specifications. Research question: How does hybrid retrieval impact factual consistency scores on table-heavy subsets of the Telco-DPR benchmark compared to text-heavy subsets?.

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 5.9/10.

3 Results

14 papers retrieved. 10 claims extracted; 4 independently verified. Quality review score: 5.9/10.

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 proposed QA system achieved an MRR of 0.68 and an accuracy of 86.2% at rank 10.	×	0.12
The QA system demonstrated a 14% improvement in accuracy on the MCQ dataset when using the GPT-4 model.	×	0.11
The DHR model achieved a Top-10 accuracy of 86.2%.	×	0.12
The proposed QA system, using the RAG model and GPT-4, achieved a 14% improvement in answer accuracy compared to a previ	✓	0.27
The retriever performance at rank 10 showed an overall accuracy of 78.0% and an MRR of 0.56.	×	0.07
The RAG+GPT-4 system achieved an overall accuracy of 86.0%.	×	0.07
The Telco-DPR dataset includes a curated 3GPP corpus in a hybrid format, combining text and tables.	✓	0.29
The dataset includes a set of synthetic question/answer pairs designed to evaluate the retrieval performance of QA syste	✓	0.30
The retrieval models evaluated include BM25, DPR, and DHR.	×	0.15
The DHR model outperforms traditional methods in retrieving relevant technical information.	✓	0.22

References

- <http://arxiv.org/abs/2410.19790v1>
- <http://arxiv.org/abs/2404.07220v2>
- <http://arxiv.org/abs/2402.12317v2>