

Dense vs. Hybrid Retrieval Accuracy on Telco-DPR Structured Data Splits

Assignee Research

June 2, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: What is the difference in retrieval accuracy between dense-only and hybrid methods when evaluated on the Telco-DPR benchmark's structured data splits. 17 claims were extracted from source literature; 5 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: Telco-DPR: A Hybrid Dataset for Evaluating Retrieval Models of 3GPP Technical Specifications. Research question: What is the difference in retrieval accuracy between dense-only and hybrid methods when evaluated on the Telco-DPR benchmark's structured data splits?.

2 Methodology

Systematic literature search across multiple databases yielded 10 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

10 papers retrieved. 17 claims extracted; 5 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
The proposed QA system achieved an MRR of 0.68 and an accuracy of 86.2% at rank 10.	×	0.11
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%.	×	0.11
The proposed QA system, using the RAG model and GPT-4, achieved a 14% improvement in answer accuracy compared to a previ	✓	0.28
The retriever performance at rank 10 for easy questions was 66.6% accuracy and 0.43 MRR.	×	0.05
The retriever performance at rank 10 for intermediate questions was 78.4% accuracy and 0.61 MRR.	×	0.05
The retriever performance at rank 10 for hard questions was 94.7% accuracy and 0.65 MRR.	×	0.05
The overall retriever performance at rank 10 was 78.0% accuracy and 0.56 MRR.	×	0.06
The QA system accuracy for RAG+GPT-4 was 87.0% for easy questions, 86.0% for intermediate questions, 84.0% for hard ques	×	0.05
The Telco-DPR dataset includes a curated 3GPP corpus in a hybrid format, combining text and tables.	✓	0.30
The dataset includes a set of synthetic question/answer pairs designed to evaluate the retrieval performance of QA syste	✓	0.31
The retrieval models evaluated include BM25, DPR, and DHR.	✓	0.16
The DHR model outperforms traditional methods in retrieving relevant technical information.	✓	0.22
The proposed QA system uses the RAG technique.	×	0.13
The average number of tokens for questions is 21 and for answers is 147.	×	0.05
The most common first question word is 'What' at 79.67%.	×	0.02
The least common first question words are 'Where' at 0.40% and 'Why' at 0.34%.	×	0.02

References

- <http://arxiv.org/abs/2410.19790v1>
- <http://arxiv.org/abs/2404.07220v2>
- <http://arxiv.org/abs/2503.01763v2>