

Query-Type Specific Retriever Ensembles Enhance Cross-Domain Robustness in Open-Domain QA

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: What is the effect of query-type specific retriever ensembles on cross-domain robustness when tested on the NaturalQuestions and TriviaQA benchmarks. 8 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Analysing the Robustness of Dual Encoders for Dense Retrieval Against Misspellings. Research question: What is the effect of query-type specific retriever ensembles on cross-domain robustness when tested on the NaturalQuestions and TriviaQA benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.8/10.

3 Results

13 papers retrieved. 8 claims extracted; 0 independently verified. Quality review score: 2.8/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
On clean questions, data augmentation, contrastive learning, and their combination do not harm the retrieval performance	×	0.13
All robustification approaches (Data Augmentation, Contrastive Learning, and Combined) perform significantly better than	×	0.13
The combined approach of data augmentation and contrastive learning achieves the highest performance among all tested me	×	0.08
Robustness of Dual Encoders deteriorates when typos are restricted to non-stopwords or discriminative utterances compare	×	0.11
The most significant performance losses occur when typos appear in discriminative utterances (lexical matches with the r	×	0.06
The proposed data augmentation combined with contrastive learning approach remains the best performing method across all	×	0.05
There is a strong positive correlation between the frequency of typoed words in the training set and retrieval performan	×	0.03
Retrieval performance drops significantly as the frequency of the typoed words in the training set decreases.	×	0.02

References

- <http://arxiv.org/abs/2404.07220v2>
- <http://arxiv.org/abs/2110.06918v3>
- <http://arxiv.org/abs/2205.02303v1>