Instruction-Following Enhances Robustness in Retrieval Models Against Adversarial Queries

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: To what extent does instruction-following capability in retrieval models improve robustness against adversarial query perturbations as measured by recall metrics on TREC Deep Learning datasets. 8 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.6/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Towards Better Instruction Following Retrieval Models. Research question: To what extent does instruction-following capability in retrieval models improve robustness against adversarial query perturbations as measured by recall metrics on TREC Deep Learning datasets?.

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 3.6/10.

3 Results

13 papers retrieved. 8 claims extracted; 0 independently verified. Quality review score: 3.6/10.

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 marginal negative sampling strategy reduces computational complexity from combinatorial to linear, i.e., O(|B| $\cdot$ |y|	×	0.02
The univariate contrastive objective in Eq. (6) may experience competition among its individual terms.	×	0.02
The multivariate objective in Eq. (7) formulates a more challenging ranking-based contrastive task by introducing a larg	×	0.04
The multivariate formulation potentially exhibits greater robustness to competition-related issues.	×	0.04
FollowIR-7B has 50.7K parameters, 5.9M training samples, and 16.1K test samples.	×	0.02
e5-base-v2 + InF-Embed achieves a score of 13.4 on the Robust04 dataset.	×	0.05
e5-large-v2 + InF-Embed achieves a score of 17.5 on the Robust04 dataset.	×	0.06
ModernBERT-base + InF-Embed achieves a score of 4.29 on the Robust04 dataset.	×	0.05

References

- http://arxiv.org/abs/2505.21439v1
- http://arxiv.org/abs/2308.12039v1
- http://arxiv.org/abs/2504.05181v2