

Adversarial Training on Synthetic Misspellings Enhances Zero-Shot Retriever Generalization

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: To what extent does adversarial training on synthetic misspelling datasets improve the zero-shot generalization of dense retrievers on out-of-domain noisy question benchmarks like TriviaQA. 9 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. 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: To what extent does adversarial training on synthetic misspelling datasets improve the zero-shot generalization of dense retrievers on out-of-domain noisy question benchmarks like TriviaQA?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.5/10.

3 Results

15 papers retrieved. 9 claims extracted; 0 independently verified. Quality review score: 3.5/10.

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
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	×	0.12
The combined approach of data augmentation and contrastive learning achieves the highest performance among all tested me	×	0.08
Robustness of Dual Encoder models deteriorates when typos are restricted to non-stopwords or discriminative utterances c	×	0.09
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	×	0.09
There is a strong positive correlation between the frequency of typoed words in the training set and retrieval performan	×	0.10
On the Natural Questions test set with random typos, the original Dual Encoder achieves an AR@5 of 49.52.	×	0.05
On the MS MARCO Dev set with typos in discriminative utterances, the original Dual Encoder achieves an MRR@10 of 10.51.	×	0.05

References

- <http://arxiv.org/abs/2403.10939v1>

- <http://arxiv.org/abs/2205.02303v1>
- <http://arxiv.org/abs/2212.09598v3>