

Robustness of Multilingual Embedding Models in Adversarial Cross-Lingual Retrieval

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: What is the robustness of multilingual embedding models in cross-lingual retrieval tasks when evaluated on adversarial or noisy queries in low-resource languages, using metrics like mean reciprocal. 8 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. 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 robustness of multilingual embedding models in cross-lingual retrieval tasks when evaluated on adversarial or noisy queries in low-resource languages, using metrics like mean reciprocal rank (MRR) and normalized discounted cumulative gain (NDCG)?.

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 3.7/10.

3 Results

15 papers retrieved. 8 claims extracted; 0 independently verified. Quality review score: 3.7/10.

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	×	0.14
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	×	0.12
The most significant performance losses occur when typos appear specifically in discriminative utterances (lexical match	×	0.03
The proposed data augmentation combined with contrastive learning approach remains the best performing method across all	×	0.08
There is a strong positive correlation between the frequency of typoed words in the training set and retrieval performan	×	0.10
Discriminative utterances identified by overlapping consecutive words between ground-truth passages and questions are ty	×	0.02

References

- <http://arxiv.org/abs/2511.19325v1>
- <http://arxiv.org/abs/2205.02303v1>
- <http://arxiv.org/abs/2301.12566v1>