

Cross-Lingual Retrieval Density and Answer Accuracy in Multilingual RAG Systems

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How does cross-lingual retrieval density impact answer accuracy in multilingual RAG systems compared to monolingual baselines on the XQuAD benchmark. 13 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Synergistic Approach for Simultaneous Optimization of Monolingual, Cross-lingual, and Multilingual Information Retrieval. Research question: How does cross-lingual retrieval density impact answer accuracy in multilingual RAG systems compared to monolingual baselines on the XQuAD benchmark?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.8/10.

3 Results

14 papers retrieved. 13 claims extracted; 1 independently verified. Quality review score: 4.8/10.

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
XQuAD-R is a question-answering dataset containing parallel questions and passages in 11 languages.	×	0.04
MLQA-R is a question-answering dataset containing parallel questions and passages in 7 languages.	×	0.04
The evaluation models were trained on datasets completely separate from XQuAD-R, MLQA-R, and MIRACL, ensuring no data sa	×	0.08
Mean Average Precision (mAP) is the metric reported for XQuAD-R and MLQA-R evaluations.	×	0.03
In the XQuAD-R dataset, there are 110 cross-lingual retrieval settings and 11 monolingual retrieval settings.	×	0.10
In the MLQA-R dataset, there are 42 cross-lingual retrieval settings and 7 monolingual retrieval settings.	×	0.10
X-X sampling strategy performs well in monolingual retrieval settings but not in cross-lingual settings.	×	0.10
X-Y sampling strategy performs well in cross-lingual retrieval settings but not in monolingual settings.	×	0.10
Hybrid batch sampling achieves the best performance in multilingual retrieval settings compared to X-X and X-Y sampling	×	0.14
When using XLM-R as initialization, Hybrid sampling achieved a monolingual (Mo.) score of 0.798, a cross-lingual (Cr.) s	×	0.07
When using LaBSE as initialization, Hybrid sampling achieved a monolingual (Mo.) score of 0.817, a cross-lingual (Cr.) s	×	0.07
Hybrid batch training substantially reduces language bias in multilingual retrieval compared to monolingual training.	✓	0.38
The proposed approach uses machine translation to generate parallel question-answer pairs across several languages, incl	×	0.09

References

- <http://arxiv.org/abs/2404.07220v2>
- <http://arxiv.org/abs/2408.10536v1>
- <http://arxiv.org/abs/2605.31171v1>