

Chunk Size Effects on Retrieval Recall and Answer Faithfulness in Religious Domain RAG Systems

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How does varying chunk size (sentence vs. paragraph) impact retrieval recall and answer faithfulness scores in RAG systems evaluated on religious domain benchmarks like QuranQA. 14 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: Reconstructing Context: Evaluating Advanced Chunking Strategies for Retrieval-Augmented Generation. Research question: How does varying chunk size (sentence vs. paragraph) impact retrieval recall and answer faithfulness scores in RAG systems evaluated on religious domain benchmarks like QuranQA?.

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 3.5/10.

3 Results

14 papers retrieved. 14 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
Fixed-Window Chunking and Semantic Chunking techniques do not differ significantly in performance when using the Jina-V3	×	0.05
Fixed-Window Chunking is easier to implement and faster than Semantic Chunking.	×	0.05
The Rank Fusion technique yields improved results when chunks are enriched with additional context from the document.	×	0.05
Adding a final reranking step is crucial for achieving consistent improvements in Rank Fusion workflows.	×	0.01
Late Chunking performs well in most cases compared to Early Chunking.	×	0.06
Late Chunking does not consistently outperform Early Chunking across all models and datasets.	×	0.05
On the NFCorpus dataset using the BGE-M3 model, the Early Chunking approach outperforms the Late Chunking approach.	×	0.06
On the MsMarco dataset using the Stella-V5 model, the Early Chunking approach outperforms the Late Chunking approach.	×	0.06
ContextualRankFusion obtains better overall results than Late Chunking on the NFCorpus subset when using the Jina-V3 mod	×	0.04
The Jina-V3 model has an MTEB Rank of 53 and a size of 572M parameters.	×	0.02
The Stella-V5 model supports a maximum token length of 131,072.	×	0.01
For the Jina-V3 model with Fixed-Window Chunks (FCC) on the TR dataset, the RFR method achieved an NDCG@10 of 0.308 comp	×	0.05
For the Stella-V5 model on the dataset in Table 4, Early Chunking achieved an NDCG@5 of 0.630 while Late Chunking achiev	×	0.04
In the comparison between Late Chunking and ContextualRankFusion, the Contextual method achieved an NDCG@5 of 0.317 vers	×	0.05

References

- <http://arxiv.org/abs/2603.25333v1>
- <http://arxiv.org/abs/2510.25621v1>
- <http://arxiv.org/abs/2504.19754v1>