SOVEREIGN: VideoRAG: Retrieval-Augmented Generation with Extreme Long-Context Videos

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 27, 2026

Abstract

Retrieval-Augmented Generation (RAG) has demonstrated remarkable success in enhancing Large Language Models (LLMs) through external knowledge integration, yet its application has primarily focused on textual content, leaving the rich domain of multi-modal video knowledge predominantly unexplored. This paper introduces VideoRAG, the first retrieval-augmented generation framework specifically designed for processing and understanding extremely long-context videos. Our core innovation lies in its dual-channel architecture that seamlessly integrates (i) graph-based textual knowledge grounding for

1 Introduction

Analysis of: VideoRAG: Retrieval-Augmented Generation with Extreme Long-Context Videos. Research goal: What is the precision drop for LLMs on HotPotQA under noisy context when scaling context window size from 32K to 128K, and does iterative retrieval with reranking mitigate this degradation more effectively across different model families?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score $<$ 6.5.

3 Results

12 papers retrieved. 6 claims extracted, 6 verified. Tribunal: 7.8/10 $\rightarrow$ APPROVE (revision_round=0). Policy: AUTO_APPROVE.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

5 Extracted Claims

Claim	Verified	Confidence
VideoRAG is the first retrieval-augmented generation framework specifically designed for processing and understanding ex	✓	0.36
VideoRAG's core innovation is a dual-channel architecture that integrates graph-based textual knowledge grounding and mu	✓	0.32
VideoRAG can process unlimited-length videos by constructing precise knowledge graphs that span multiple videos.	✓	0.30
The LongerVideos benchmark comprises over 160 videos totaling 134+ hours across lecture, documentary, and entertainment	✓	0.25
VideoRAG demonstrates substantial performance compared to existing RAG alternatives and long video understanding methods	✓	0.31
The source code of VideoRAG implementation and the benchmark dataset are openly available at https://github.com/HKUDS/Vi	✓	0.29

References

- http://arxiv.org/abs/2510.22344v1
- http://arxiv.org/abs/2502.01549v1
- http://arxiv.org/abs/2403.09832v1