

CORTEX: A MULTI-MODAL RETRIEVAL-AUGMENTED GENERATION FRAMEWORK WITH KNOWLEDGE GRAPH AUGMENTATION AND INTELLIGENT ANSWER SYNTHESIS

Shubham Chaurasia, Pranjali Mishra, Pradyumn Prajapati, Prince Yadav & Sheelu Singh

Computer Science and Engineering, Axis Institute of Technology and Management, Kanpur, U.P, India

ABSTRACT

If you've worked with any real-world document collection, you'll know the problem — data comes in from everywhere, PDFs, spreadsheets, audio recordings, video, and somehow a system has to make sense of all of it together. Most RAG systems we looked at were basically text-only, and even then they chunked documents in ways that made retrieval worse rather than better. Cross-document relationships? Almost entirely ignored. We built CORTEX to tackle these issues, and this paper describes what we built, how it works, and what actually happened when we tested it. It handles nine file formats, uses embedding-based chunking to split at topic boundaries rather than arbitrary character counts, and enriches each chunk with context before embedding. At query time, three retrieval signals — Qdrant vector search, BM25, and knowledge graph traversal — run in parallel and their results are fused using RRF. None of the three alone gave us what we wanted; the combination was what mattered. We also added a CRAG-style self-evaluation loop that checks if the answer is actually supported by what was retrieved. When it isn't — which happened more than we expected in early tests — it decomposes the query and tries again. The platform works with both locally-hosted LLMs through Ollama and cloud-based providers like Anthropic and OpenAI. Responses can be delivered in five formats: cited answers, hierarchical summaries, flashcards, timelines, and interactive knowledge graph views. It also automatically detects contradictions across uploaded documents. Results show meaningful gains over single-method baselines, though we acknowledge the evaluation setup has limitations we discuss later.

KEYWORDS: *Retrieval-Augmented Generation, Knowledge Graphs, Semantic Chunking, Hybrid Retrieval, Corrective RAG, Multi-Modal Processing, Reciprocal Rank Fusion, Vector Databases*

Article History

Received: 24 Apr 2026 | Revised: 25 Apr 2026 | Accepted: 28 Apr 2026
