



72-Hour Technical Challenge

Multimodal Enterprise RAG – Leveraging Knowledge Graphs and Hybrid Search

Objective

Design and implement a modular prototype of an **Enterprise Retrieval-Augmented Generation (RAG)** system that supports text, image, audio, and video ingestion, builds a **searchable knowledge graph**, and enables **hybrid search** using, keyword and vector retrieval (Graph RAG System).

You are expected to **begin with evals** to define success criteria and to structure your architecture accordingly.

Enterprise Approach Highlights

Before building any ingestion or pipeline logic, define your **evaluation framework**:

- What constitutes a "correct" response?
- What types of queries will you support (e.g., factual, lookup, reasoning)?
- What metrics will you track (e.g., hallucination rate, latency, accuracy)?
- How will you fail gracefully?

Structure your work around a **modular and scalable pipeline**, including:

- **Input validation**
- **Query triage and rewriting**
- **Agent-based retrieval orchestration**

- **Hybrid Search: Structured Graph Traversal + Keyword Filtering + Semantic Vector Retrieval**
 - **Answer generation and post-processing**
-

Challenge Scope

Build a multimodal RAG assistant that:

- Ingests at least three of the following modalities: text, image, audio, video
 - Extracts entities and relationships
 - Constructs a searchable knowledge graph (e.g., Neo4j or similar)
 - Build a searchable Vector Database like Qdrant or Weaviate in parallel with a sophisticated ingestion pipeline
 - Powers a hybrid search pipeline for **fast and reliable access to domain-specific knowledge**
-

Requirements

1. Evaluation-First Pipeline Design

- Define a minimal test suite using DeepEval or similar
- Clearly document:
 - Query types (lookup, summarization, semantic linkages)
 - Evaluation goals: retrieval quality, hallucination control, latency
- Include functional unit tests for each module

2. Data Ingestion and Preprocessing

- Accept: .pdf, .txt, .jpg/.png, .mp3/.mp4
- Modal-specific logic:
 - OCR/captioning
 - Transcription
 - Frame extraction and tagging for video

- Enrich all outputs with metadata and domain tags

3. Entity & Relationship Extraction

- Use LLMs to extract structured information
- Cross-modal linking of the same entity (e.g., “John Smith” in PDF + transcript)
- Generate or infer schema for graph database

6. User Interface / Demo

- UI or notebook should support:
 - Uploading new files
 - Typing natural language queries
 - Viewing answers with optional graph exploration
 - Log evaluation output for each query
-

Bonus Features

- Scene detection for video
 - Sentiment detection from text/audio
 - Topic-based reranking of results
 - Real-time feedback for query improvement
 - Security-aware design (query restrictions, access control)
-

Evaluation Criteria

- **Enterprise fit:** eval-first mindset, modularity, and clear architecture
- **Precision and relevance:** does the system retrieve the right context across modalities?
- **Latency:** fast, efficient retrieval and generation
- **Reliability:** graceful failure handling and consistent outputs
- **Maintainability:** clear logic, good documentation, testing

Resources & Tips

There's no strict requirement for the type of dataset or topic. Suggested tools and sample sets:

Frameworks

- AutoGen: <https://github.com/microsoft/autogen>
- CrewAI: <https://github.com/joaomdmoura/crewAI>
- DeepEval: <https://github.com/confident-ai/deepeval>
- Arize Phoenix: <https://github.com/Arize-ai/phoenix>
- LangChain: <https://www.langchain.com/>
- LlamaIndex: <https://www.llamaindex.ai/>
- Chonkie: <https://github.com/trychonkie/chonkie>

Multimodal Tools

- LLaVA: <https://llava-vl.github.io/>

Vector Databases

- Qdrant, Milvus, Weaviate

Sample Datasets

- DocVQA: <https://docvqa.github.io/>
- PubLayNet: <https://github.com/ibm-aur-nlp/PubLayNet>
- LAION-400M: <https://laion.ai/blog/laion-400-open-dataset/>

Submission

Within 72 hours, submit:

- GitHub repository or zip file
- Short demo video (3–10 minutes)
- Local setup instructions
- Optional: your evaluation report and reflections on architecture decisions