Case Study: Multimodal RAG + Enterprise Document Analysis

Theme

Multimodal RAG + Enterprise document analysis

Duration

3 days

Context

Your platform must analyze unstructured and structured documents (PDFs, tables, screenshots, logs, emails, invoices), combine them semantically, and generate dependable summaries, insights, and workflows. The solution must run inside a secure Kubernetes cluster, though Kubernetes itself is not tested here. A Notebook (e.g., Google Colab) is recommended for demonstration.

Task

- Design and partially implement a proof-of-concept RAG pipeline that achieves the following:
- Ingests a small set of PDFs (3–5 provided by us: a report, a product specification, a table-heavy document).
- Extracts text, tables, and images (optional but highly valued).
- Builds a hybrid RAG index / Knowledge Base.
- Implements retrieval methods that search indexes, explain why chunks were retrieved, synthesize answers with citations, and ensure only relevant chunks/images/information are used.
- Produces a summary + structured JSON output including key findings, extracted numerical data, risk flags, and domain-specific insights.

Deliverables

- An architecture diagram (Mermaid, draw.io, or a hand sketch).

- A Jupyter/Python notebook or small repo with:
    - ingestion pipeline
    - chunking strategy
    - hybrid RAG
    - agent logic
    - evaluation examples

- A 5–10 minute video or README covering:
    - Chunking design for tables
    - Retrieval decision process
    - Hallucination prevention
    - Scaling approach for millions of documents
    - Adjustments for on-prem enterprise clients