



## Problem Statement

**Title:** Improving Accuracy and Reducing Hallucinations in Retrieval-Augmented Generation (RAG) Using Multi-Agent Orchestration and Selective Multimodal Parsing

### **Abstract:**

Retrieval-Augmented Generation (RAG) systems aim to enhance response quality and factual grounding by leveraging external documents. However, current RAG implementations often struggle with hallucinations due to incomplete or misinterpreted context, and they largely ignore or naively convert non-textual content such as images. In real-world applications, documents frequently contain visual elements (charts, figures, diagrams) that are critical for context but may not translate meaningfully through standard OCR or captioning methods.

This project proposes an **Agentic RAG framework** that uses **multi-agent orchestration** to improve accuracy, reduce hallucinations, and handle multimodal document content without forcibly converting images to text. Instead of extracting raw textual descriptions from images, the system will **identify, classify, and intelligently route image-containing sections** to **specialized agents** that can flag their importance, defer human review, or augment the textual context through external metadata or descriptions.

By breaking the RAG workflow into coordinated agents — such as document segmenters, retrievers, verifiers, image-aware routers, and response generators — the system gains flexibility, modularity, and robustness. This approach not only improves retrieval quality but also helps the generator avoid hallucinations by recognizing the limits of the available context, especially when images contain key information.

### **Key Research Questions:**

1. How can hallucinations in RAG be mitigated when critical context exists in non-textual form?
2. What are effective strategies for handling images in documents without converting them to raw text?
3. How can agent-based orchestration be applied to coordinate retrieval, validation, and generation steps?
4. How should RAG models be designed to acknowledge knowledge gaps (e.g., due to inaccessible images) rather than hallucinate plausible but false responses?

## Objectives:

- Build a modular, multi-agent RAG pipeline with image-awareness that avoids naive text conversion.
- Design an agent that detects and classifies the relevance of images within documents.
- Implement fallback and deferral mechanisms (e.g., ask user, flag uncertainty) for unprocessable image content.
- Incorporate hallucination detection and mitigation strategies such as agent consensus, verifier checks, or abstention triggers.
- Evaluate the system on a dataset of multimodal documents with a mix of text and images, focusing on factual accuracy, uncertainty calibration, and reduction in hallucinations.

## Submission Guidelines:

### 1. Technical Architecture Report (PDF or DOCX):

- a. Executive Summary.
- b. Problem Statement and Motivation.
- c. Detailed System Architecture (with diagrams, agent workflows, and interaction logic).
- d. Methodology and Implementation Details.
- e. Evaluation Methodology, Metrics, and Results.
- f. Conclusion and Recommendations.
- g. References and Appendices (if applicable).

### 2. Codebase Submission (GitHub Repository):

- a. Well-structured codebase with clear modular organization.
  - b. Readme file including:
    - Project description.
    - Setup instructions.
    - Execution steps and usage guide.
    - Sample dataset and expected output formats.
- Clear comments and documentation in the code.
  - Test cases and evaluation scripts (if applicable).

Submission Link: [Link](#)

Deadline: 11:59pm, 17th May, 2025