Hiring Task: AI-Based UPSC Answer Sheet Evaluation with RAG

🎯 Objective

Design and build an **AI-powered automated grading workflow** that evaluates handwritten UPSC answers using **vision and language models**, and incorporates a **RAG (Retrieval-Augmented Generation)** system grounded in our official SOP Evaluation document.

📌 Task Overview

Step 1: Input Handling

- Accept input as **PDFs or images of handwritten answer sheets**.

- These documents contain:

    - A visible **question**

    - The **student's answer**, structured across **Introduction, Body, Conclusion**

    - **Marking scheme** (e.g., "15 marks") embedded in the question text

Step 2: Workflow Automation

You must build an **end-to-end pipeline** that includes:

a. OCR Extraction

- Extract handwritten content using a robust OCR system (e.g., **Google Vision API**, **PaddleOCR**, **Amazon Textract**, or vision-LMs like GPT-4o).

- Segment extracted text into structured fields: `question`, `name`, `introduction`, `body`, `conclusion`.

b. RAG-based Evaluation

- Implement a **Retrieval-Augmented Generation system**:

  - The retriever must reference a knowledge base built from the shared PDF:
    `SOP_Evaluation_Test Series_Mentorship_Hindi.pdf`

  - You must chunk, embed, and store this SOP in a **vector database** like Pinecone, Weaviate, or Qdrant.

  - When evaluating a student's response, the generator (LLM) should:

    - Retrieve relevant SOP snippets (in Hindi) per section (Intro, Body, Conclusion, etc.)

    - Use them as grounding context while generating feedback and assigning marks.

c. Evaluation Criteria

Evaluate each section based on:

- **Content Adequacy**

- **Structure (IBC format)**

- **Depth of Argument & Analysis**

- **Use of Examples / Data**

- **Presentation**

- **Language and Grammar**

- **Alignment with the question's directive keyword** (e.g., "Critically examine", "Discuss")

Step 3: Feedback Generation

The final report must follow this structured format:

Dear Student,
Total Marks: X.X

Introduction

[3-line comment aligned with SOP criteria]

Body

• [Comment 1]
 • [Comment 2]
 • [Comment 3]
 Marks Awarded: X.X

Conclusion

• [Comment 1]
 • [Comment 2]
 • [Comment 3]
 Marks Awarded: X.X

Presentation

Strengths:
 • [Point 1]
 • [Point 2]

Improvements:
 • [Point 1]
 • [Point 2]
 • [Point 3]
 Marks Awarded: X.X

Strengths

• You have a fair understanding of the topics.
 • Your contextual understanding is really appreciable.

• You have used the Introduction – Body – Conclusion approach.
• Your handwriting is neat and legible.

All the Best! Keep improving and striving for excellence.

🛠️ Tooling Requirements

Candidates can use modern tooling such as:

- **LLMs**: GPT-4o, Claude 3, Gemini, Mistral, LLaMA3

- **OCR**: PaddleOCR, Tesseract, Vision APIs

- **RAG Components**:

    - **Embeddings**: Cohere, OpenAI, SentenceTransformers

    - **Vector DB**: Pinecone, Qdrant, Weaviate

    - **Frameworks**: LangChain, LlamaIndex

- **Frontend**: Streamlit, Gradio, or FastAPI

- **Optional Enhancements**:

    - Re-rankers

    - Section-specific chunking

    - Diagram detection and markup

✅ Deliverables

1. GitHub link (or ZIP) containing:

   - Code

   - Preprocessed SOP

   - Chunking and embedding strategy

   - Inference flow

2. Example Input: 1 sample answer sheet (you may simulate)

3. Example Output: Corresponding evaluation report

4. README with:

   - Tech stack

   - RAG architecture diagram

   - Evaluation pipeline explanation

🌟 Bonus Points

- Hosted demo on Hugging Face Spaces / Streamlit Cloud

- Hindi language handling and alignment with SOP tone

- Vision-LLM integration (e.g., GPT-4o)

📅 Timeline

Submit your task within **2 days** from assignment.