

Take-Home Assignment: Scalable Question Generation System

Objective

Build a minimum viable product (MVP) that generates high-quality, multiple-choice questions from large text documents. Your solution can be inspired by the concepts in the "[Savaal](#)" research paper, focusing on a creative and scalable design rather than an exact replication.

Core Requirements

1. **Input:** The system must ingest and process PDF or plain text documents that you are provided with.
2. **Scalability:** Design a pipeline to efficiently handle large documents. Your design should account for potential issues like memory limits, API costs, and context loss.
3. **Question Generation:** The system must produce conceptual multiple-choice questions. Each question should have:
 - One correct answer grounded in the source text.
 - 3-4 plausible but incorrect "distractor" options.
4. **Output:** The final output must be a single JSON file containing the list of generated questions, their choices, and the correct answer for each.

Bonus Features (Optional)

- **Quality Control:** Implement a mechanism to automatically score, rank, or filter questions to ensure high quality.
- **Difficulty Management:** Add a feature to generate questions based on difficulty (e.g., easy, medium, hard). *Hint: Researching Bloom's Taxonomy may be useful.*

Tech Stack

- **Language:** Python
- **LLM API:** Your choice (e.g., OpenAI, Anthropic, Gemini).

Deliverables

1. **Jupyter/Colab Notebook (.ipynb):** Submit a single, well-commented notebook with setup instructions and markdown cells explaining your design choices. **Do not include API keys in your submission.**
2. **JSON Output File:** The JSON file generated by running your notebook on the given set of documents.
3. **Video Demo (max 3 minutes):** A brief recording that demonstrates your system, explains its architecture, and shows the output.

Evaluation Criteria

- **Functionality & Quality:** Does the system work? How relevant and challenging are the questions?
- **Scalability & Design:** How thoughtful is your technical approach to handling large inputs?
- **Code Quality:** Is the code clean, organized, and well-documented?
- **Communication:** How clearly do you explain your work in the notebook and video?

Submission Details

- Please submit your completed assignment within **5 business days** of receiving this assignment to be considered for the interview step.
- Send your final submission in a .zip file to **siyuz8561@gmail.com**.