AI/ML Engineer Technical Assessment: RAG Chatbot with STT/TTS

Overview

Build a multimodal chatbot application that demonstrates proficiency in RAG (Retrieval Augmented Generation), speech processing, and a user interface.

Requirements

1. **Frontend Interface** (Choose one)
   - Streamlit (recommended for rapid prototyping)
   - Flask with HTML/CSS/JavaScript
   - FastAPI with frontend framework
2. **Large Language Model Integration**
   - OpenAI GPT (GPT-3.5/GPT-4)
   - Anthropic Claude
   - Open-source models (Ollama, Hugging Face)
   - Google Gemini
   - Any other LLM API
3. **RAG Implementation**
   - Document ingestion and preprocessing
   - Vector embeddings generation
   - Vector database integration
   - Semantic search and retrieval
   - Context-aware response generation
4. **Speech-to-Text (STT)**
   - Audio input recording
   - Speech recognition (OpenAI Whisper, Google Speech-to-Text, etc.)
   - Audio format handling (WAV, MP3, etc.)

Technical Specifications

Input Method

- Audio recording button with real-time feedback

Output Method

- Text with playback audio

RAG Knowledge Base

- Use the given document
    - **Big data analytics—A review of data-mining models for small.pdf**

Deliverables

1. Source Code

- Complete application code
- Requirements.txt or environment.yml
- Clear project structure
- README with setup instructions

2. Video Submission

- **Video Demonstration** (3-5 minutes)
    - Feature walkthrough
    - Text and audio input examples
    - RAG functionality showcase
    - Technical explanation in English

3. Deployment (Bonus)

- Docker containerization
- Cloud deployment (Heroku, AWS, GCP)
- Live demo URL

Submission Requirements

- Upload your Project Files
    - Zip your project folder
    - Name the zip file: AI_[YOUR_FULL_NAME].zip
        - Example: AI_John_Smith.zip
    - Include setup instructions in README.md
- Upload the Demo Video into Google Drive
    - Name the video file: AI_[YOUR_FULL_NAME].mp4
    - Record a video (3-5 minutes) demonstrating all implemented features
    - Walk through each feature and the implementation in English
    - Explain any bonus features implemented
- Delivery
    - Send via email with subject line: "AI/ML Engineer- [Your Name]"
    - Attach your Google Drive link with the video & codes
    - Include any additional notes or assumptions in the email body