AI Agent Architecture Document

Overview

The AI Mock Interview Coach is a modular system designed to automate interview preparation tasks such as resume parsing, personalized question generation, audio transcription of user responses, and constructive feedback generation. The system leverages specialized AI models integrated into a structured pipeline, providing users with an intuitive interface and seamless experience.

Components

1. Resume Parsing Agent

- Purpose: Extract structured data (full name, email, technical skills, etc.) from a resume.
- Model Used: Ollama LLM integrated via LangChain.
- Reason for Choice: Ollama enables flexible prompt-based interaction, while LangChain provides an easy way to structure prompts and handle responses.
- Input: Raw resume text (TXT file).
- Output: JSON object with extracted fields.

2. Question Generation Agent

- Purpose: Generate personalized technical interview questions based on extracted skills.
- Model Used: Fine-tuned Llama 3 using PEFT (LoRA).
- Reason for Choice: Llama is efficient for instruction-based tasks, and LoRA allows cost-effective fine-tuning with small datasets.
- Input: List of technical skills.
- Output: List of interview questions.

3. User Interaction Agent (Speech-to-Text)

- Purpose: Transcribe spoken candidate answers into text.
- Model Used: OpenAI Whisper.
- Reason for Choice: Whisper provides robust, accurate transcription across multiple audio formats and qualities.
- Input: Audio file (MP3, WAV).
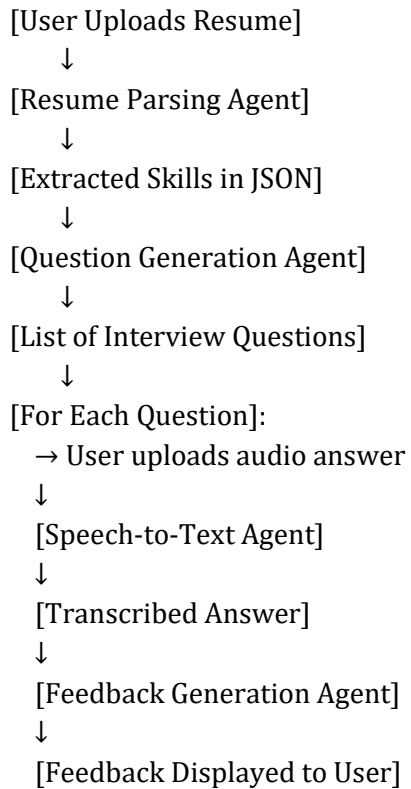- Output: Transcribed answer text.

4. Feedback Generation Agent

- Purpose: Generate detailed, constructive feedback for the candidate's answer.
- Model Used: Fine-tuned Llama 3 (same as question generator).
- Reason for Choice: Enables structured, encouraging feedback via prompt-driven generation.
- Input: Interview question + Transcribed answer.
- Output: Feedback text.

5. Streamlit Frontend

Provides user interaction for:
- Resume upload
- Audio upload
- Displaying questions, transcribed answers, and feedback
- Session state management

Interaction Flow Diagram

[User Uploads Resume]
     ↓
[Resume Parsing Agent]
     ↓
[Extracted Skills in JSON]
     ↓
[Question Generation Agent]
     ↓
[List of Interview Questions]
     ↓
[For Each Question]:
   → User uploads audio answer
   ↓
   [Speech-to-Text Agent]
   ↓
   [Transcribed Answer]
   ↓
   [Feedback Generation Agent]
   ↓
   [Feedback Displayed to User]

This architecture ensures modularity, scalability, and clear separation of concerns between components.