AI Interview Assistant Prompt Library Documentation

Overview

This project implements an AI Interview Assistant using LangChain, demonstrating the use of different language models (GPT-4 and Ollama's llama3) for conducting mock interviews. The system is designed to adapt its interview style based on the candidate's experience level, the type of skills being assessed, and the desired tone of the conversation.

Project Structure

- test.ipynb: Main Jupyter notebook containing the implementation
- README.md: This documentation file

Features

1. Customizable Interview Parameters

The system accepts three main parameters to customize each interview session:
- Level: beginner or advanced
- Categorization: technical, creative, or analytical
- Tone: formal, casual, or funny

2. Interview Components

Each interview session includes:
1. Introduction
2. 'Tell me about yourself' question
3. Follow-up questions
4. Detailed feedback
5. Closing remarks

3. Model Integration

The project demonstrates integration with two different LLM models:
- OpenAI's GPT-4
- Ollama's llama3 (local model)

Model Comparison

Prompt and Output Analysis

Test Case 1: Beginner Technical Interview (Formal Tone)

Aspect	GPT-4	Ollama (llama3)
Introduction	Formal and structured	Friendly and supportive
Initial Question Style	Structured and context-based	Direct and simple
Follow-up Questions	Technical, project-based, learning approach	Not shown
Feedback Structure	Comprehensive and detailed	Basic acknowledgment
Response Length	~400-500 words	~100-150 words

Test Case 2: Advanced Creative Interview (Formal Tone)

Aspect	GPT-4	Ollama (llama3)
Introduction	Formal and sophisticated	Creative and welcoming
Initial Question Style	Sophisticated, vision-based	General with creative focus
Follow-up Questions	Portfolio, innovation, industry insight	Not shown
Feedback Structure	Strategic, multi-dimensional	Basic notes
Response Length	~600-700 words	~150-200 words

Test Case 3: Beginner Analytical Interview (Casual Tone)

Aspect	GPT-4	Ollama (llama3)
Introduction	[Not shown]	Relaxed, friendly opening
Initial Question Style	[Not shown]	Casual and encouraging
Tone Adaptation	[Not shown]	Effective casual tone
Response Length	[Not shown]	~100-150 words

Key Differences in Implementation

Feature	GPT-4	Ollama (llama3)
Response Structure	Hierarchical, well-organized	Linear, conversational
Question Complexity	Adaptive by level/category	Simpler structure
Context Retention	Maintains throughout	Focus on immediate interaction
Tone Adaptation	Precise to formal/casual/funny	Basic adaptation
Technical Depth	Level-based adjustment	Uniform depth
Feedback Style	Structured, rich feedback	General, short feedback
Resource Usage	API-based, higher latency	Local, faster response
Cost Implications	Pay per token	Free local model

Best Practices

1. Model Selection:
- Use GPT-4 for: Advanced scenarios, detailed feedback, complex roles
- Use Ollama for: Basic practice, quick feedback, local usage

2. Parameter Configuration:
- Match 'level' with experience
- Align 'categorization' with role
- Adjust 'tone' based on company culture

Future Improvements

1. Add more industry-specific categories
2. Implement structured output parsing
3. Add multi-turn conversation capability
4. Include question banks for industries
5. Develop performance evaluation metrics