1. Project Overview

The goal of this project is to automate Excel skill evaluation during candidate interviews using an AI-powered conversational agent. This addresses the current bottleneck of manual assessments that consume analyst time and cause inconsistent results. The system simulates an interviewer, asks structured Excel questions, evaluates candidate responses using an LLM, and provides a detailed feedback report.

2. Technology Stack and Justifications

Component	Choice	Justification
Frontend	Streamlit	Fast prototyping, great for building interactive apps, allows easy state handling and rapid deployment. Ideal for PoC.
Backend Logic	Python + REST API (Groq)	Python offers simple HTTP integration and session management. Calling LLM via HTTP ensures flexibility.
LLM	LLaMA3-8B via Groq API	LLaMA3 offers high-quality reasoning at lower cost. Groq delivers extremely fast inference (latency < 1s), making user interaction smooth.
Evaluation	Prompt-based LLM Scoring	Lightweight and flexible without needing a training dataset upfront.
Hosting	Streamlit Community Cloud	Free, simple, and publicly accessible — perfect for demonstration and early deployment.
Environment Variables	python-dotenv	Used to manage secrets like the Groq API key securely.

3. System Flow

1. App loads and welcomes the user.
2. Questions are either uploaded via `.txt` or selected from defaults.
3. Each question is shown one by one.
4. User submits their answer.
5. LLaMA3 is prompted to evaluate the response and return:
   - A score (0–5)
   - Feedback
   - Improvement suggestions
6. Answers and feedback are stored.
7. At the end, a transcript is saved and shown to the user.

4. Cold Start Strategy

This system is designed to operate without a pre-existing dataset. Here's how the cold start is handled:

Bootstrapping without Data:

- The system uses a pre-defined set of Excel questions.

- Evaluation is prompt-driven via a general-purpose LLM (LLaMA3).
- Prompts are crafted to follow strict output format and evaluation rubric.

Improvement Over Time:

- Analysts can review AI feedback to validate or adjust scores.
  This forms a labeled dataset that can later be used to:
  - Fine-tune a lightweight evaluation model.
  - Improve prompt design using real-world patterns.

User Feedback Collection:

- A thumbs up/down system can be added to assess the usefulness of AI feedback.
- Helps track evaluation quality at scale.

5. Future Enhancements

- Support feedback export via email or download.
- Multi-language support using LLaMA's multilingual capabilities.