

README: Admissions Counselor Voice AI Agent

Submitted by: Ankit Kumar Chandresh Singh

1. Prompt & Conversation Flow

The agent's logic is governed by a detailed system prompt that defines its persona ("Alex," a professional admissions counselor) and its core rules.

- **Goal-Oriented:** The primary goal is to assist students by providing accurate course information. A secondary goal is to gather the caller's name and course interest.
- **Instruction-Driven:** The prompt uses strict instructions, such as mandating the use of the `get_course_info` tool for all data queries and enforcing an exact-phrase response for out-of-scope questions to prevent hallucination.
- **Safety Guardrails:** The agent is explicitly forbidden from mentioning any real-world university details, ensuring confidentiality and focus.
- **Conversation Flow:** The designed flow is logical and user-centric: **Greet → Identify Intent → Use Tool to Answer → Handle Follow-ups / Out-of-Scope → Conclude.**

2. Tools & Technical Stack

This agent was built using a stack of reliable and performant tools.

- **Platform:** **Vapi** (for real-time conversation management).
- **Language Model (LLM):** **Llama 3 8B**, accessed via the **Together AI** API.
- **Backend Server:** **Python (Flask)** server hosted on **Replit** to serve the primary tool.
- **Speech-to-Text:** **Gladia** was explicitly configured for robust transcription.

3. Data Storage & Retrieval

To ensure low latency and simplicity, the course data table was stored as an in-memory JSON object on the Replit server rather than in an external database.

- **Retrieval Method:** A single tool, `get_course_info`, is exposed to the LLM. This function takes `course_name` or `level` (e.g., "bachelor," "master") as arguments, searches the JSON data, and returns the precise course details for the agent to articulate.

4. Edge Cases Handled

The agent was specifically designed to be resilient to common user interaction challenges.

- **Ambiguity:** If a user asks for a generic course like "BCom" (for which 3 versions exist in the data), the agent asks for clarification instead of guessing.
- **Broad Filtering:** For general questions like "Do you have any master's programs?", the agent correctly uses the level parameter in its tool to filter the data and provide a comprehensive list.
- **Out-of-Scope Queries:** Any question not answerable by the data tool (e.g., about campus life) triggers a pre-defined, safe response, ensuring the agent never provides false information.

A Note on Performance: This agent is deployed entirely on free-tier services. As a result, you may experience a minor "cold start" latency on the first query if the agent has been idle. This is a known characteristic of the serverless architecture. Subsequent interactions will be significantly faster.