Assignment: Design an End-to-End AI Voice Assistance Pipeline

Objective:

Design a pipeline that takes a voice query command, converts it into text, uses a Large Language Model (LLM) to generate a response, and then converts the output text back into speech. The system should have low latency, Voice Activity Detection (VAD), restrict the output to 2 sentences, and allow for tunable parameters such as pitch, male/female voice, and speed.

Step 1:

Voice-to-Text Conversion - In this step explain how you will convert voice input (microphone or audio file) to text. Use any open source Speech2Text Model preferably -

Whisper

- https://github.com/openai/whisper
- https://github.com/ggerganov/whisper.cpp
- https://github.com/SYSTRAN/faster-whisper

Use a pre-trained English model (e.g., en-US) with the following settings:

- Sampling rate: 16 kHz
- Audio channel count: 1 (mono)
- VAD threshold: 0.5 (to detect voice activity and ignore silence)

Step 2:

Text Input into LLM - In this step you are required to input the query into LLM in text format

- Use Hugging Face Transformers with a pre-trained model
- Model: Choose an LLM model that is suitable for your specific application (e.g., llama, Mistral, Mixtral, Phi2 etc)
- The converted Speech to text output will be used as a query for above LLM

Step 3:

Text-to-Speech Conversion - In this step you are required to convert text into speech

- Use any Text to Speech open source models such as -
    - https://github.com/rany2/edge-tts
    - https://huggingface.co/microsoft/speecht5_tts
    - https://huggingface.co/suno/bark
    - https://huggingface.co/parler-tts/parler-tts-large-v1

The output of the LLM from the previous step 2 should be converted into speech (file .mp3 or .wav)

Additional Requirements:

1. Latency: Suggest how to minimize the latency under (< 500 ms) using WRTC ?
2. VAD: Implement VAD to detect voice activity and ignore silence.
3. Output Restriction: Restrict the output response to a maximum of 2 sentences.
4. Tunable Parameters: Allow for tunable parameters such as:
    - Pitch: adjust the pitch of the synthesized speech
    - Male/Female Voice: choose between different voices (e.g., Joanna or Samantha)
    - Speed: adjust the speed of the synthesized speech

Evaluation Criteria:

Documentation - Detailed explanation of the E2E implementation and architecture

Demo - Working Demo (recording file)

Code Quality - Quality of the code written to implement the pipeline.

Submission Requirements:

Please submit a PDF document containing:

1. A detailed description of your solution, including the choice of models, libraries, and parameters used.
2. Code snippets (in Python) demonstrating how you implemented each step of the pipeline.
3. Any relevant documentation or diagrams that support your implementation.

Reference

https://github.com/huggingface/speech-to-speech

https://github.com/livekit/agents/

Disclaimer

Please note that this assignment is for informational purposes only, and we do not expect you to provide fully functional or working code as part of your submission.

All intellectual property rights (including but not limited to the code, concepts, and ideas) contained in your submission remain with the candidate. We respect these rights and will not use any submitted materials without permission from the candidate.

If you are selected for an interview, we may request a code walkthrough or coding assessment as part of the face-to-face interview process. This is intended to be a collaborative discussion where you can explain your design choices, implementation details, and problem-solving strategies in more depth.

Your submission should focus on demonstrating your understanding of the assignment requirements, design decisions, and technical skills rather than providing fully functional code. We are looking for evidence of your thought process, problem-solving abilities, and communication skills, which will be further evaluated during the interview process.

By submitting this assignment, you acknowledge that you have read and understood these terms and conditions. If you have any questions or concerns, please do not hesitate to reach out to us.

Thank you for your interest in our company, and we look forward to reviewing your submission!