AI Audio & Text Summarization App

This Python script provides a simple web interface using Gradio to transcribe English audio using OpenAI's Whisper model and summarize the resulting text using a language model served locally via LM Studio. It allows processing both uploaded/recorded audio and direct text input.

Features

- **Audio Transcription:** Transcribes English audio files (uploaded or recorded) using various sizes of the Whisper model.
- **Text Summarization:** Summarizes the transcribed audio text or directly provided text using a language model hosted in LM Studio.
- **Local Models:** Utilizes locally downloaded Whisper models and a local LM Studio server for processing.
- **Gradio Interface:** Provides an easy-to-use web interface for interaction.
- **File Saving:** Saves original audio, transcript, and summary to local directories.
- **Configurable:** Allows changing LM Studio settings, Whisper model path, and default model size.

Requirements

- Python 3.7 or higher
- pip (Python package installer)
- FFmpeg (required by Whisper for audio processing)
- LM Studio (Download from https://lmstudio.ai/)
- A compatible language model loaded in LM Studio (e.g., deepseek-r1-distill-qwen-7b as configured in the script).

Installation

1. **Save the code:** Save the provided Python code as a .py file (e.g., main.py).
    - *Note:* If you downloaded the dl project zip folder, the main script is located at DL_PROJECT\ai_voice_text_summarizer\main.py.
2. **Install Python dependencies:** Open your terminal or command prompt, navigate to the directory where you saved the file (or the ai_voice_text_summarizer directory if using the dl project zip), and run:
   pip install gradio openai-whisper numpy scipy

   (Note: openai-whisper is the correct package name for the Whisper model library.)

3. **Install FFmpeg:**
    - **Windows:** Download from https://ffmpeg.org/download.html and add the bin directory to your system's PATH.
    - **macOS (using Homebrew):** brew install ffmpeg
    - **Linux (using apt):** sudo apt update && sudo apt install ffmpeg
4. **Set up LM Studio:**
    - Download and install LM Studio.
    - Launch LM Studio and download a compatible language model (the script is configured for deepseek-r1-distill-qwen-7b, but you can change LM_STUDIO_MODEL_ID in the script).
    - Go to the "Local Inference Server" tab (usually the 3rd icon from the left).
    - Load your chosen model into the server.
    - Start the server. Ensure it is running and accessible at the base URL specified in the script (http://127.0.0.1:1234/v1 by default).

Usage

1. **Run the script:** Open your terminal or command prompt, navigate to the script's directory (DL_PROJECT\ai_voice_text_summarizer\ if using the dl project zip), and run:
   python main.py

2. The script will print messages about initializing models and starting the Gradio interface. It will provide a local URL (usually http://127.0.0.1:7860).
3. **Access the web interface:** Open a web browser and go to the URL provided by the script.
4. **Process Audio:**
    - Go to the "Process Audio" tab.
    - Upload an English audio file or record audio using your microphone.
    - Select the desired Whisper model size from the dropdown. Larger models are more accurate but require more resources and take longer to load (especially the first time).
    - Click the "Transcribe and Summarize Audio" button.
    - The transcript and summary will appear in the respective textboxes. A download link for a .txt file containing both will also appear.
    - Processing status updates will be shown in the "Processing Status" box.
5. **Process Text:**
    - Go to the "Process Text" tab.
    - Paste or type English text into the "Enter English Text" textbox.

    - Click the "Summarize Text" button.
    - The summary will appear in the "Summary" textbox. A download link for a .txt file containing the original text and summary will also appear.
    - Processing status updates will be shown in the "Processing Status" box.
6. **Stop the application:** Close the terminal window where the script is running, or press Ctrl+C in the terminal.

Project Structure

The image provided shows that this application (ai_voice_text_summarizer) is part of a larger DL_PROJECT. The main application code and its generated files (audio, models, summaries) reside within the ai_voice_text_summarizer directory. Other directories at the DL_PROJECT level contain related code and resources, such as fine-tuning scripts and project documentation.

File Structure (within ai_voice_text_summarizer)

Assuming you are running the script from the DL_PROJECT\ai_voice_text_summarizer\ directory, the script will create the following directories in that location:

- audio/: Stores uploaded and recorded audio files with timestamps.
- summaries/: Stores saved summary and transcript/original text files with timestamps.
- models/: Stores downloaded Whisper model files.

Related Projects

This application is part of a larger set of projects within the DL_PROJECT structure. The code for fine-tuning the models potentially used with this application can be found at the following locations:

- **Whisper Model Fine-tuning:** D:\DL_PROJECT\whisper_model_finetuning
- **DeepSeek Fine-tuning:** D:\DL_PROJECT\deepseek_finetuning

Troubleshooting

- **LM Studio Connection Error:** Ensure LM Studio is running, the server is started, and the correct model is loaded. Verify the LM_STUDIO_BASE_URL in the script matches your LM Studio server address.
- **Whisper Model Loading Error:** Check the console output for specific errors. Common issues include missing FFmpeg, insufficient RAM/VRAM for the selected model size, or permission issues writing to the models directory.
- **No Speech Detected:** Whisper might return "[No speech detected]" if the audio quality is poor, contains significant background noise, or is indeed silent.

- **Gradio Interface Not Loading:** The default port (7860) might be in use. Try running the script with a different port, e.g., app.launch(share=False, server_name="0.0.0.0", server_port=7861).