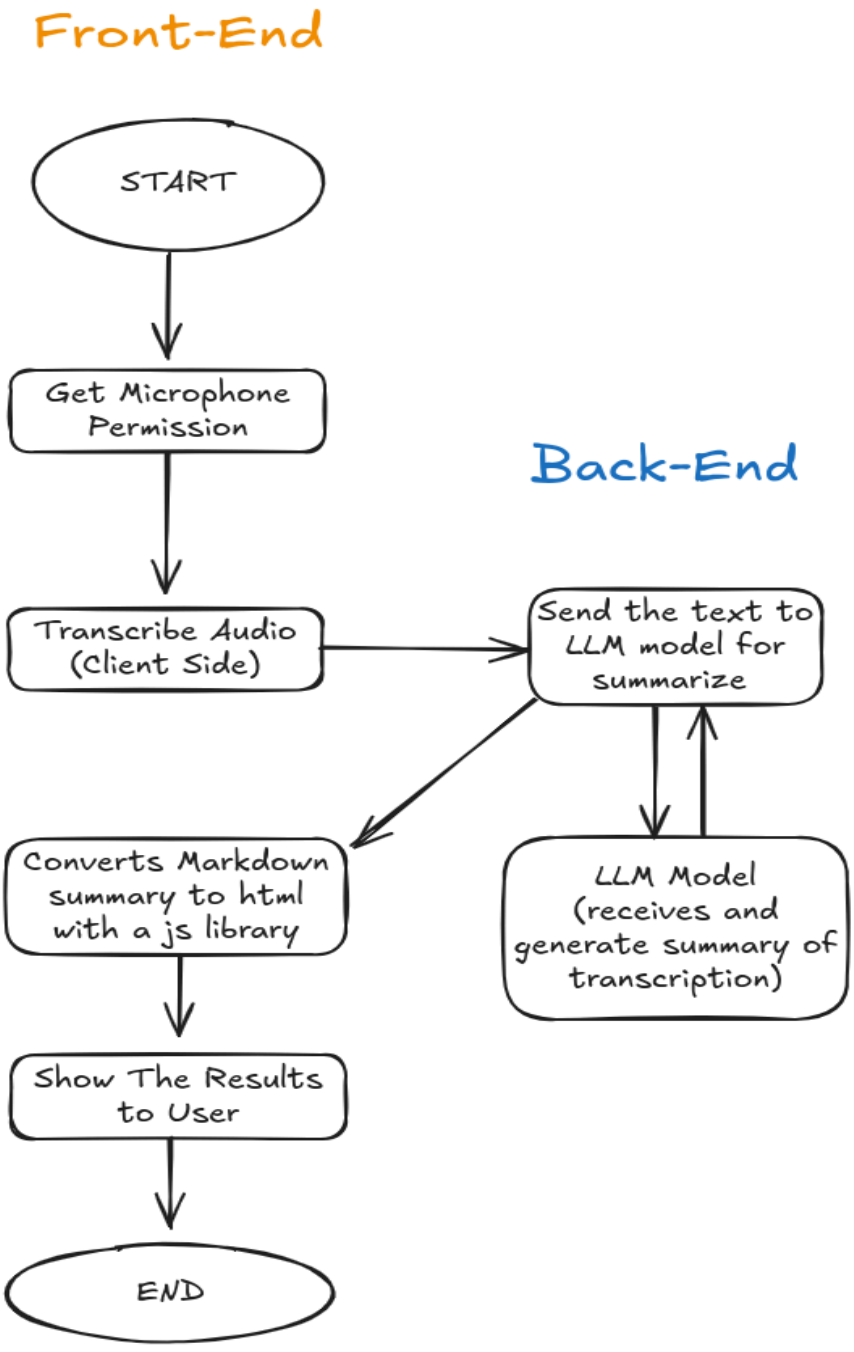
Architecture



Front-End

1. **Start**
   The user initiates the process (for example, by clicking a “Record” button).

2. **Request Microphone Access**
   The app asks the user’s browser for permission to use the microphone. If permission is denied, the flow stops and the user is informed; if granted, it proceeds.

3. **Capture and Transcribe Speech**
   While the user speaks, the browser listens and turns their words into a raw text transcript in real time. As speech is recognized, chunks of text become available for the next stage.

Back-End

4. **Send Transcript for Summarization**
   Once the spoken content is fully captured, the front end sends that raw transcript to your server as a simple text package.

5. **Generate a Structured Summary**
   Your server forwards the transcript to a chosen large language model. The model processes the text, extracts the main ideas, and returns a concise, well-organized summary—often formatted in a lightweight markup style.

Front-End (Again)

6. **Format the Summary for Display**
   The returned summary (in its original markup form) is converted into user-friendly, styled text so it looks clean and readable in the browser.

7. **Present Both Transcript and Summary**
   The interface shows two panels (or sections): one displaying exactly what was

transcribed word-for-word, the other showing the model's distilled version. This side-by-side view lets users verify the raw text and quickly absorb the key points.

8. **End / Ready for Next Run**
   With both views visible, the user can review, copy, or start a new recording. The system returns to its idle state, ready to repeat the process.