AI DOCTOR 2.0
(Vision and Voice)

Virtual Doctor that can see, listen and respond like a real doctor using AI

PROJECT LAYOUT

The project is divided into four main phases:

- ***PHASE 1 - Setup the brain of Doctor***

This is where we build the core intelligence using a multimodal LLM. We use the GROQ API to process input, and we also convert images into the format needed for the model to analyze.

- ***Phase 2: Setup the voice of the patient***

Here, we let users speak. We use an audio recorder with tools like ffmpeg and portaudio to capture the voice, and then transcribe it using a speech-to-text model.

- ***Phase 3: Setup the voice of the doctor***

This is where the doctor speaks back. We use text-to-speech tools like gTTS and ElevenLabs to convert the model's response into audio.

- ***Phase 4: User interface for the VoiceBot***

Finally, we create an interactive UI using Gradio so users can talk to the AI doctor in real time.

Tools and Technologies

- ***Groq*** – For fast AI model inference.
- ***OpenAI Whisper*** – This is one of the best open-source tools for transcribing speech to text.
- ***Llama 3 Vision*** – An open-source model by Meta that can understand images.
- ***gTTS and ElevenLabs*** – For generating the doctor's voice.
- ***Gradio*** – To build the user interface.
- And, of course, we coded everything using ***Python in VS Code***.

Technical Architecture