AI DOCTOR 2.0

(Vision and Voice)

Medical Chatbot with MultiModal LLM

PROJECT LAYOUT

Phase 1–Setup the brain of the Doctor (Multimodal LLM)

- Setup GROQ API key
- Convert image to required format
- Setup Multimodal LLM

Phase 2–Setup voice of the patient

- Setup Audio recorder (ffmpeg & portaudio)
- Setup Speech to text–STT–model for transcription

Phase 3–Setup voice of the Doctor

- Setup Text to Speech–TTS–model (gTTS & ElevenLabs)
- Use Model for Text output to Voice

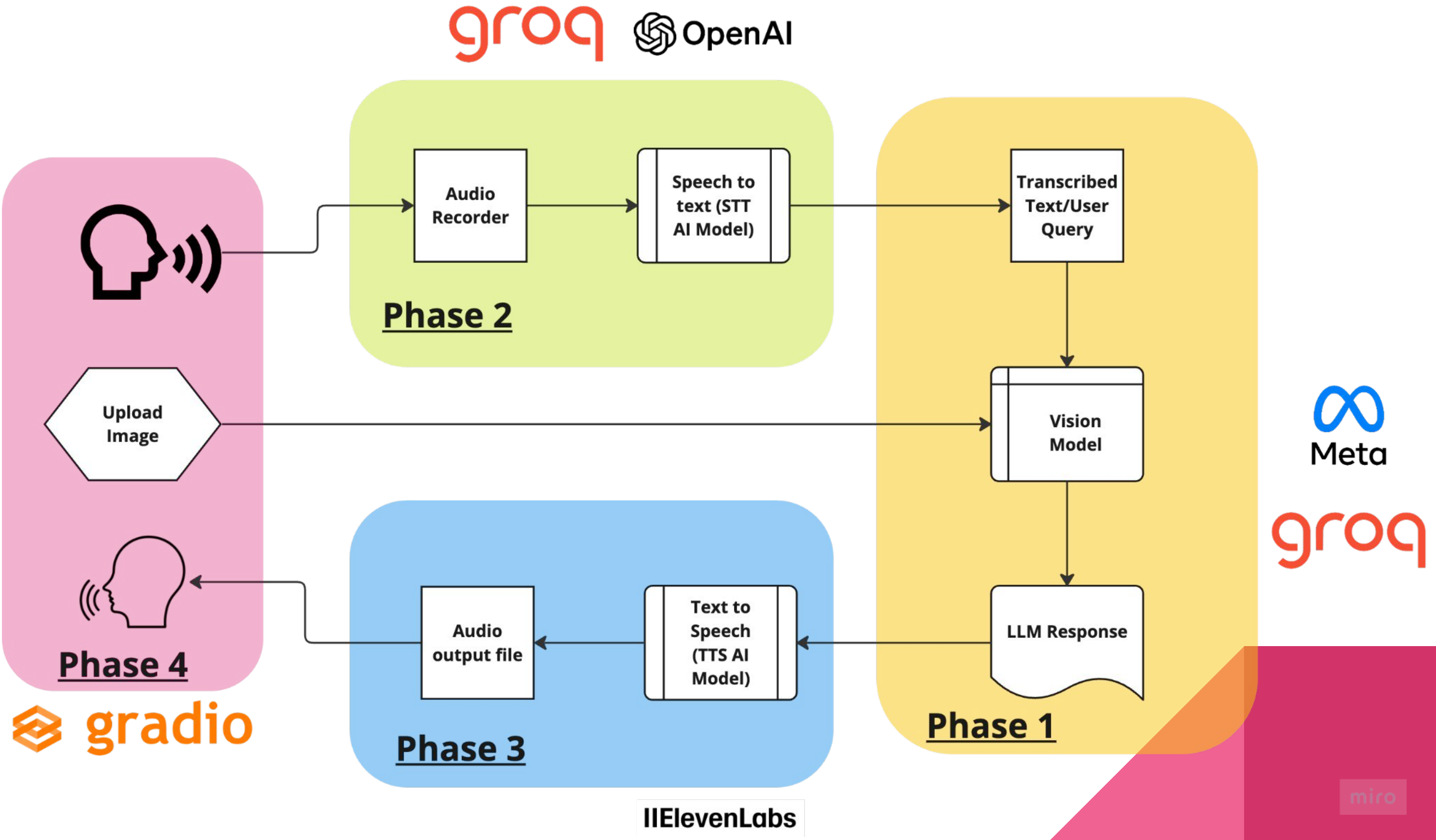
Phase 4–Setup UI for the VoiceBot

- VoiceBot UI with Gradio

TOOLS AND TECHNOLOGIES

- Groq for AI Inference
- OpenAI Whisper (Best open source model for Transcription)
- Llama 3 Vision (Open source by Meta)
- gTTS & ElevenLabs (Speech to Text)
- Gradio for UI
- Python
- VS Code

TECHNICAL ARCHITECTURE



IMPROVEMENT POTENTIAL/NEXT STEPS

- Use state-of-the art LLMs, especially for vision (Paid LLMs)
- Finetune vision model on Medical images
- Add multilingual capabilities