AI Doctor 2.0: Vision and Voice

A medical chatbot powered by a multimodal LLM, enabling voice, text, and vision-based healthcare assistance.

By Adarsh Gandhi

Project Layout: Building the AI Doctor

1

Phase 1: AI Brain

Set up Multimodal LLM with GROQ API.

2

Phase 2: Patient Voice

Integrate audio recorder and STT model.

3

Phase 3: Doctor Voice

Implement TTS model for voice output.

4

Phase 4: VoiceBot UI

Develop UI with Gradio.

Tools and Technologies

Groq

AI Inference

Whisper

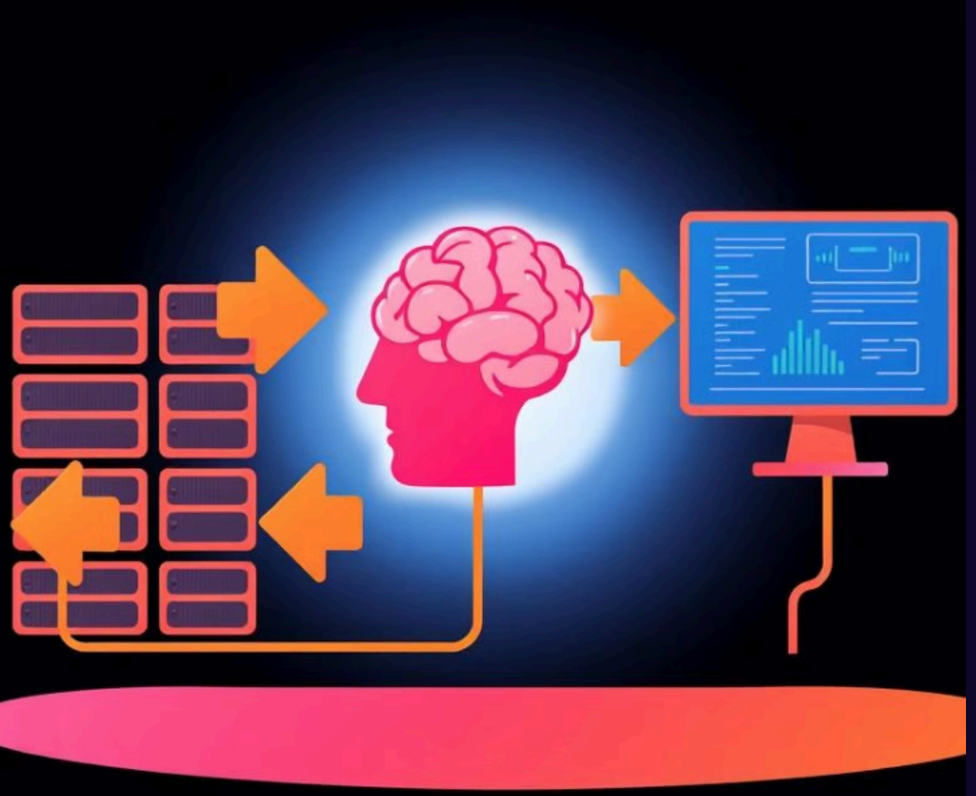
Transcription

Llama 3 Vision

Vision

Python

Language



Technical Architecture Overview

AI Doctor 2.0 architecture for multimodal interaction.

Phase 1: Setting up the AI Brain

1 **GROQ API Key**

Setup GROQ API key for AI inference.

2 **Image Conversion**

Convert images to required format.

3 **Multimodal LLM**

Setup Multimodal LLM for processing.

Phase 2: Patient Voice Input

Audio Recorder

Setup audio recorder (ffmpeg & portaudio).

Speech to Text

Setup STT model for transcription.

Phase 3: Doctor Voice Output

Text to Speech

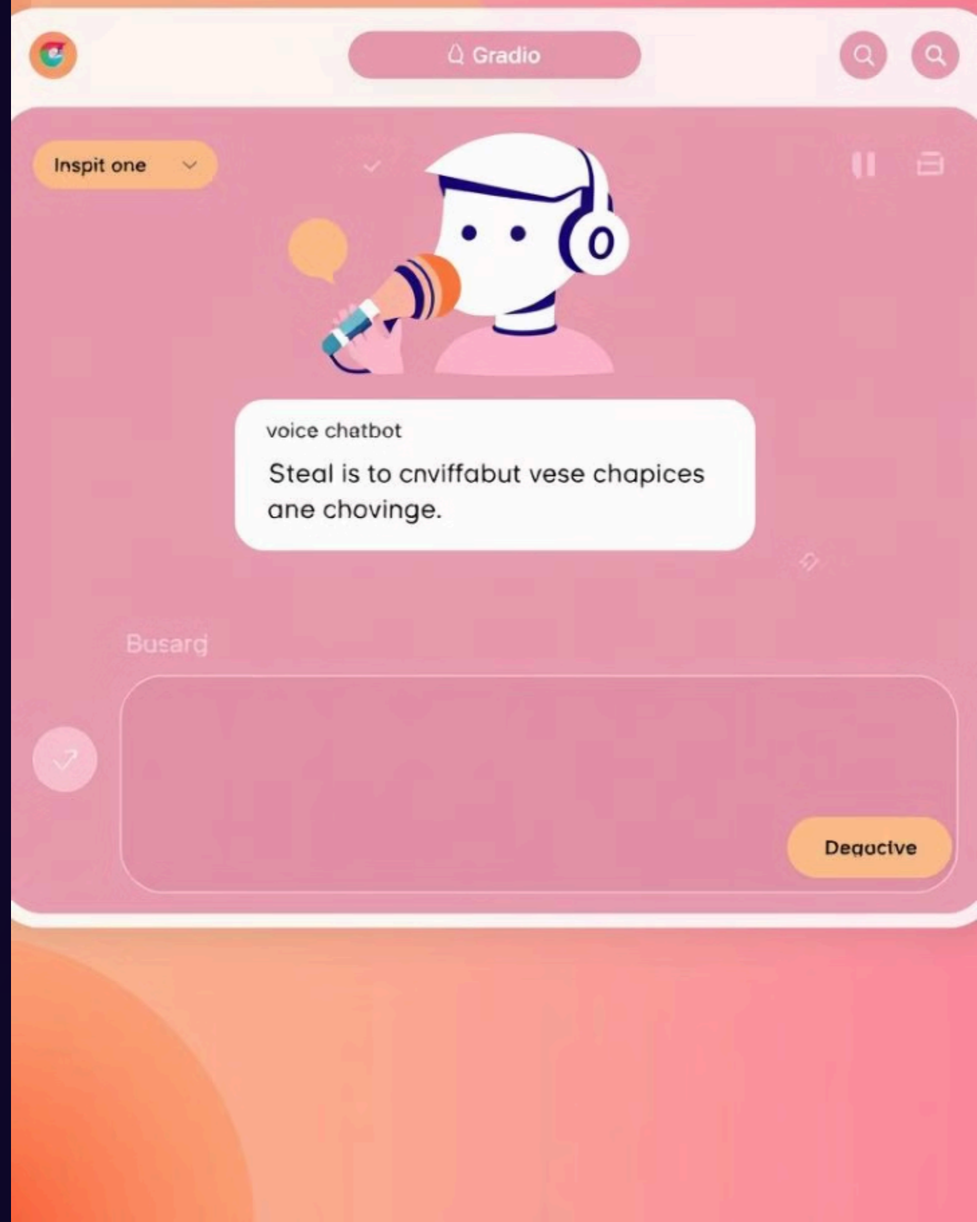
We're setting up a Text-to-Speech (TTS) model, using gTTS & ElevenLabs, to convert text into audible speech.

Voice Output

We'll utilize this model for generating voice output directly from the text, creating the doctor's voice.

Phase 4: VoiceBot UI with Gradio

A VoiceBot UI built with Gradio, enabling seamless and interactive user engagement through voice commands.



Improvement Potential/Next Steps

1

State-of-the-art LLMs

Use paid LLMs for better vision.

2

Finetune Vision Model

Finetune on medical images.

3

Multilingual Capabilities

Add multilingual support.

Thank You

Thank you all for your time and attention today. I'm excited to continue this journey and see how we can further enhance the patient experience through advanced AI and voice capabilities. Please let me know if you have any other questions - I'm happy to discuss further.