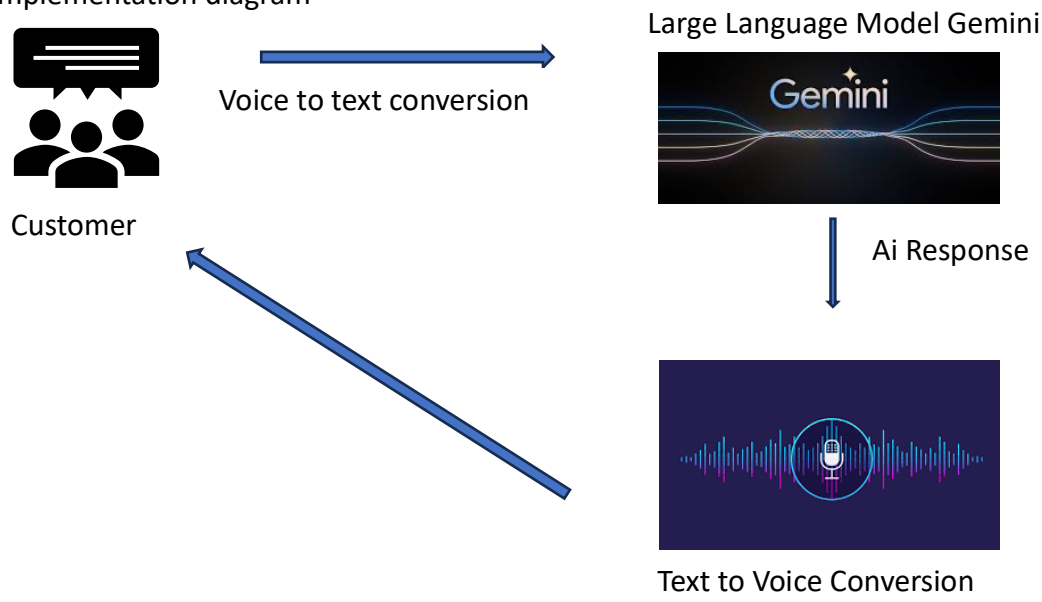
AI Voice Assistant

Implementation diagram



I have Created streamlit-based web application which allows users to record their voice, transcribes the audio into text, generates a response using Google Gemini's model and then converts the AI's response back into speech.

Step 1 Voice-to-Text Conversion –
I have used :

a) Streamlit - A framework to create web apps in Python.

b) Audio_recorder_streamlit - which captures audio input from the user within the Streamlit app.

c) Faster_whisper - used to convert for Audio transcription.

Step 2 Text Input into LLM -
I have used Gemni ai model which input the query into LLM in text format .

a) Generativeai – which Interacts with Google's Gemini model.

Step 3 Text-to-Speech Conversion -
Here model response is converted to speech and played back to the user. I have used :

a) Google Text-to-Speech (gTTS) - Converts text responses back into speech.
b) Pygame - to plays the audio files.

Github project Link