Vision AI

Name: Subhanshu
College: Indian Institute of Technology (Indian School of Mines) Dhanbad
Branch: Electronics and Communication Engineering
Year of study: 3rd Year
College Admission Number: 19JE0824
Email: subhanshu1604@ece.iitism.ac.in

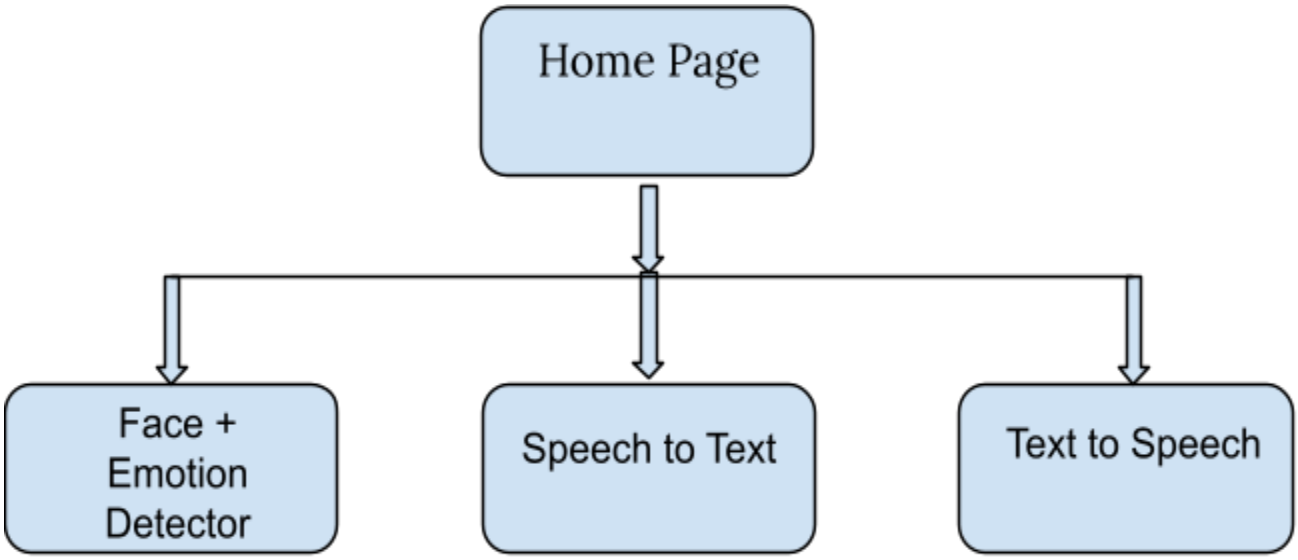
Project Aim:

To Develop both the frontend and backend of computer vision based AI systems.

Purpose of my website:

A hosted platform for the users to use the benefits of Computer vision, which in this case includes Face Detection and emotion analysis system, Speech to text and Text to speech converter.

Site Map:



Feature and Future:

The target audience can be categorized based on the following 3 use cases:

- The **face-detection and emotion analysis**
  – detects face and emotion based on the facial expression of a person analyzing the view taken from the camera of the device.
  – It can be used as a proctoring mechanism during online tests or examinations. It can also be used as a social media tool to make interactive predictions of a person's mood, based on facial expressions.
- The **speech to text**
  – helps in converting a given speech to text. The webpage contains controls to save the generated text, for future references
  – Its functionalities can be used by people having difficulty in typing or by someone who is not a fast typist.
- The **text-to-speech**
  – helps in converting a given chunk of text to speech. The webpage hosts controls to Play, Pause and Stop the generated speech.
  – on the other hand can be used by users with poor vision to hear large chunks of text without actually straining their eyes.

Tech-Stack and Implementation Details:

Frontend:

HTML-5, CSS-3, Bootstrap5, JavaScript (ES6)
use public API “Face API”

Backend:

Node.js (with express.js framework) for backend
Basic routing implemented with express.js

Responsive:

Yes (a mobile-first web-application)
Created with bootstrap-5 components and icons

Implementation of AI functionalities

AI functionalities were **implemented in Front-End**

Reason:

The website is mainly used on the client side to predict various computer vision related parameters.

Features like the face-detection, speech-to-text and text-to-speech view the need of **real-time processing of data by the ML models**. If such a task of prediction was implemented in the backend then it may lead to scalability issues, when a number of client applications are making http requests to the same

backend server.

Hence, the AI models were made to first load to the client side, and once loaded, they can use the client-side system resources to make predictions without latency.

Note: The AI models are mainly in the form of lightweight jsons, and include simple javascript functions which do not require so much system resources. Hence, client-system requirements are never a bottleneck to this design.

Role of Backend in the Website:

- The backend was made solely for routing purposes.
- It was also made expecting a future prospect of:
    - Including passport authentication for different logins based on subscription.
    - Connecting to mongoDB, for storing and saving the generated speech and texts