**Code and Documentation for the Project**

**Code**

import gradio as gr

import numpy as np

import whisper

def transcribe(audio):

    # First- we select the model for whisper AI

    # Here we are using base model because it is fast but it dose not have good accuracy in non famous languages

    model=whisper.load\_model("base")

    #Second- we load audio in our database and pad/trim audio to fit in 30 seconds

    audio = whisper.load\_audio(audio)

    audio = whisper.pad\_or\_trim(audio)

    # Third- make log-Mel spectrogram and move to the same device as the model

    mel = whisper.log\_mel\_spectrogram(audio).to(model.device)

    # Fourth- we find the probability and find out the spoken language

    \_, probs = model.detect\_language(mel)

    print(f"Detected language: {max(probs, key=probs.get)}")

    # Fifth- now we decode the audio and convert it into text

    options = whisper.DecodingOptions()

    result = whisper.decode(model, mel, options)

    return result.text

# Now we build a interface by the help of Gradio web UI

gr.Interface(

    title = 'Computer Vision Project By Using Whisper AI With The Help Of Gradio Web UI',

    fn=transcribe,

    inputs=[

        gr.Audio(sources="microphone", type="filepath")

    ],

    outputs=[

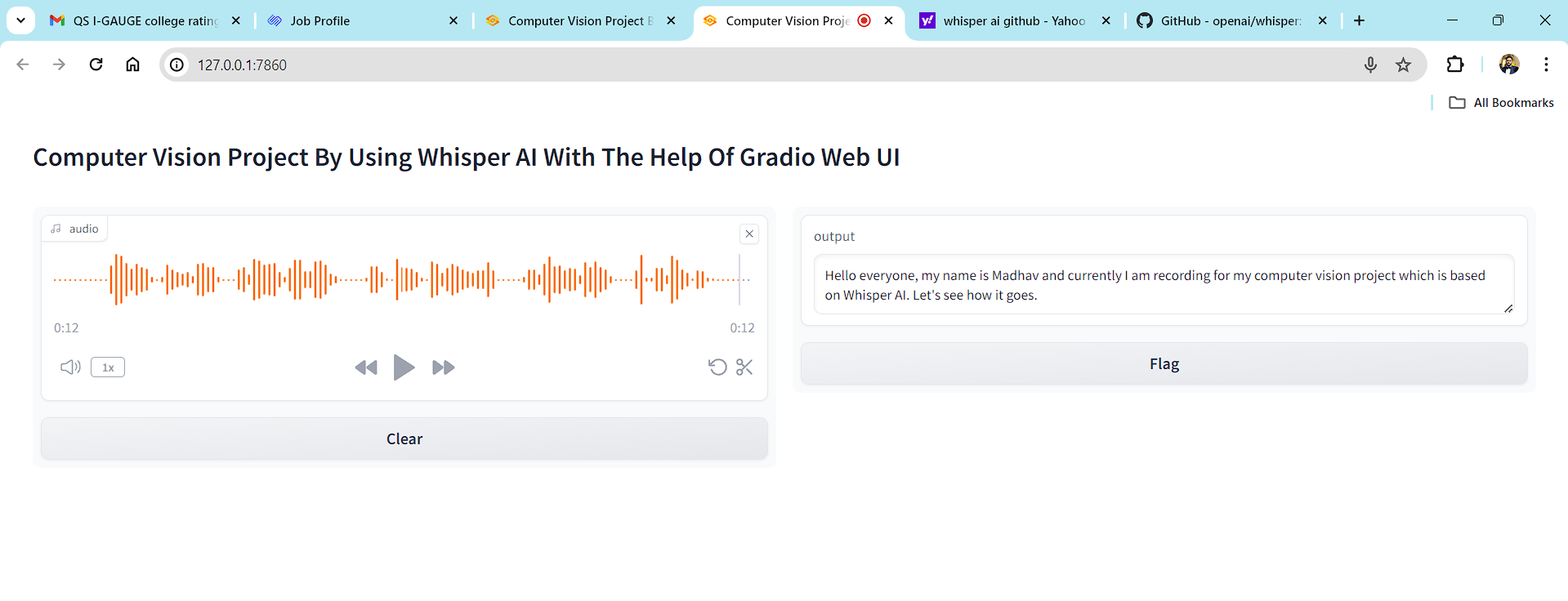
        "textbox"

    ],

    live=True).launch(share=True)

#The End

**Output**

****

**Documentation**

**Introduction:**The Multilingual Speech Recognition Model developed for this project aims to enable speech recognition across multiple languages without the need for extensive training. Leveraging a pre-trained multilingual speech recognition model, such as Multilingual Whisper, this system allows for the transcribing of speech in various languages, facilitating tasks such as translation and summarization using Whisper AI**Setup and Dependencies:**

* Ensure that Python is installed on your system.
* Install the required libraries like Whisper, Gradio, FFMPEG by running “pip install gradio”.

**Usage:**

* Run the provided code script, which includes the necessary functions for transcribing speech.
* Utilize the Gradio web UI interface to interact with the model:

**Input:** Audio files or microphone input.**Output:** Transcribed text.

**Transcription Process:**

1. User has to upload the audio file from device then pre-trained multilingual speech recognition model (e.g. Whisper) start its work task.
2. Whisper load the audio file and start processing to fit the model's input requirements (e.g., padding or trimming).
3. It will convert the audio into a log-Mel spectrogram for analysis.
4. Use the model to detect the language spoken in the audio and transcribe it into text.
5. Output the transcribed text for further processing or analysis.

**Evaluation and Quality Assurance:**

* Assess the accuracy of the transcribed text by comparing it with ground truth data across various languages.
* Ensure the clarity and completeness of the project's report, providing detailed insights into the methodology, evaluation metrics, and results.
* Conduct thorough testing of the code to verify its functionality and adherence to best practices in software development.

**Conclusion:**This Multilingual Speech Recognition Model which is built by using whisper, represents a significant advancement in enabling multilingual capabilities. By converting audio video into a accurate text, this system opens up new possibilities for cross-linguistic communication and information processing.