i want a hugging face model which can be finetuned on .mp3 files and create me a speech to text model on it give me compllete python code at once

and use the saved file to create audio when text is inputted

give me complete code at once

restructure the code like this

pretrained model and tokenizer

download and save then run

text to speech

then

use file path recordings

which contains all mp3 files

finetue the model on htis

save new model and

run text to speech again on finetuned model

and use matplotlib to identify difference between them

1. **Pretrained Model and Tokenizer**: You would need to download a pretrained model and tokenizer from Hugging Face. For speech recognition, you might want to consider models like “facebook/wav2vec2-base-960h”.
2. **Download and Save**: Once you have chosen your model, you can download and save it using the from\_pretrained method in Hugging Face Transformers.
3. **Text to Speech**: For text-to-speech, you would need a different model. Hugging Face provides TTS models like “t5-base”.
4. **Use File Path Recordings**: You would need to load your .mp3 files and convert them into a format that your model can understand. This usually involves converting the audio to a specific sample rate and possibly converting stereo audio to mono.
5. **Finetune the Model**: Finetuning the model involves training the model on your specific dataset. You would need to set up a training loop, define your loss function (likely the CTC loss for speech recognition), and then train your model.
6. **Save New Model**: After training, you can save your model using the save\_model method in Hugging Face Transformers.
7. **Run Text to Speech Again on Finetuned Model**: You can then use your finetuned model for inference.
8. **Use Matplotlib to Identify Difference Between Them**: To compare the original and finetuned models, you could run them on the same inputs and use matplotlib to plot the differences in their outputs.