Synthesising a YouTube channel

Download all the audio files from all the videos from a YouTube channel and using inference convert them into text. Then train a language model on the data to generate a fake version of the YouTuber. This can be done using Whisper from OpenAi for the speech to text part, then for the text generation part that is where the main body of my research will occur. While building this out I would like to be able to publish a general platform that would be able to automatically download parse and train a model on any channel / set of audio files with the option for someone to build their own model for training on the data set.