**Project: MIT (Music-Image Transfer)**

**Content:** input a (piece of) music with a prompt, output an image according to the audio and the requirements, and a description according to the music

* Some examples of expected model

1. 音游曲绘
2. 歌曲专辑封面
3. 演唱会海报

**Our target users:**

1. Composer / singer
2. Nowadays AI supports cheap music creations (e.g. suno), so everyone can make use of it!
3. Some hearing impaired are not able to appreciate the beauty of music

For common composers, generating a good album cover may be expensive (need to contact an expert painter and also time-consuming)

Benefit them by making this thing much easier (or at least they can draw inspirations from them)

**Challenges: (You need to think of more ideas here)**

* We need to combine the audio with user input text
* Many models in Hugging face targets audio-to-text transfer, i.e. speech recognition

So doing something with pure music (retrieving information from it) is difficult!

* Supporting long audio is challenging for some models
* Also, recognizing words in songs may be more difficult than in normal speeches
* Hard to make the model understand the emotion of the music
* Lack of labeled data (e.g. BGM with labeled emotion information)
* Adding text to image may be extremely difficult!

**How do we plan to solve these:**

* For songs with lyrics, pick a proper model to recognize the words (maybe with some fine-tuning) and add them to user’s prompt, since the lyrics can indeed represent a lot of information of this song, then send the prompt to an LLM to generate corresponding key words, then feed these words into an image-generation model
* Also, retrieve information from the background music using designed tools, such as BPM, volume, which might represent the emotion of the music.
* Or use NN to extract information from music
* When text is needed in the images, separate text with the image generated and add the text externally

**Resource**

12 nvidia 4090

5 million tokens

Vpn flow amount

Data: Music with emotion label, both with and without lyrics

(less homework on intro2sys)