MAIS 202 - PROJECT DELIVERABLE 2

1. Problem statement:

To design a model that takes a list of songs from a user as input and generates an art piece representing the names of the songs, the lyrics, and the genre. The information about the songs will be retrieved from the Spotify API and Genius. Then the processing is done using a text to image model and the art is then generated using GAN.

2. Data Preprocessing:

https://colab.research.google.com/drive/10gMMfPS9ACa8fEnDAKaz7IK4U1dMTmnM?usp=sharing

https://towardsdatascience.com/linking-images-and-text-with-openai-clip-abb4bdf5dbd2

The data about the songs in the playlist will be retrieved from the Spotify API and Genius. The features for our dataset will be song title, lyrics, energy, tempo, mode, loudness, speechiness, valence, instrumentalness, liveness,danceability, and key. These features will be scrapped from Spotify API and Genius. The chorus feature will be scrapped from the lyrics of the song.

3. Machine learning model:

After the data from the songs has been processed and the primary features have been picked, we can now use these ‘keywords’ as our basis for the art generation model. After reviewing the resources we defined online, it was determined that using a GAN from scratch would be difficult compared to the resources that we currently have since it uses high computational power and usually would need GPUs to facilitate the training. Luckily, there exist models online that we can use to our advantage to simplify the process and make it more accessible.

The framework:

The framework in the link above uses VQGAN + CLIP to generate images from text. We are yet to decide whether we will be doing our own implementation of the same model, or just using the features that serve our purposes for the scope of the project directly from the framework.

4. Preliminary results:

So far the model is producing errors running on the computer’s current dependencies so we are yet to verify the accuracy using testing and integration of the data processing.

5. Next steps:

For the next steps, we would need to verify the functionality of the VQGAN + CLIP model in generating images from text. After that, we would be able to link the two processes of fetching data and processing it from the Spotify API to the model. That way, we can now generate our training and testing data for the generation of album art on the basis of the input.