**Melodic Machines:**

[**Github Link**](https://github.com/ZahFay/CSCI1470-Melodic-Machines)

**Title**: Melodic Machines

**Who**:

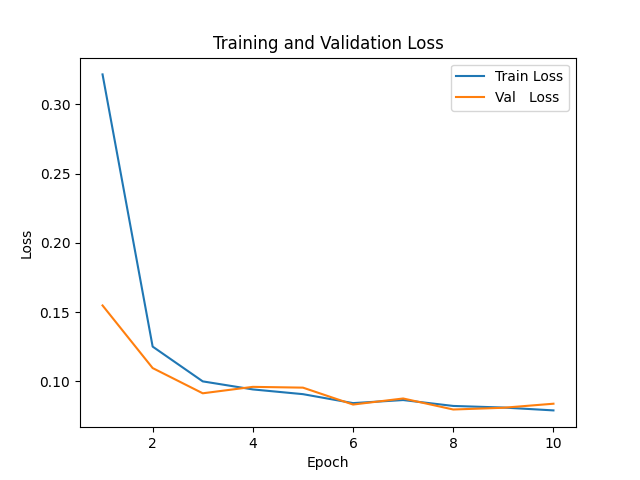
Qien Lin (qlin23)

Alzahra Fayie (afayie)

Liang Man-Fang (mliang23)

Zetao Wu (zwu119)

* **Introduction**:
  + Music creation is both an art and a science: crafting lyrics that feel “authentic” to a particular artist and producing audio that captures their signature sound can require extensive domain expertise and manual effort. Recent advances in deep generative models—transformers for text and diffusion models for images—open the door to automating these creative tasks.
  + In this project, Melodic Machines, we explore a two‑stage framework for artist‑conditioned music generation. First, a transformer‑based lyric generator takes as input an artist’s name and produces novel lyrics in their stylistic voice. Second, a stable‑diffusion model operates in the spectrogram domain to synthesize new audio textures that mimic the artist’s sonic palette. By coupling these modules, our system can output both the words and the sound of a “new” song in a target artist’s style, streamlining creative workflows and offering insights into the interplay between lyrical content and audio style.
* **Challenges**: What has been the hardest part of the project you’ve encountered so far?
  + Our biggest challenge so far was producing meaningful results with both our models. We have managed to get a really good loss value for our stable diffusion model of about 0.08 while training with Oscar; however, the generated spectrograms are still very noisy.



* + The second biggest issue with our model was the amount of significant preprocessing that we needed to do on our data. Our dataset from the FMA was originally a set of 30-second audio clips with their respective metadata, but we first clipped the audio to ten seconds, converted them to spectrograms, and converted the spectrograms into pickle files so that the model trains on matrix representations of the spectrograms, making the training faster.
* **Insights**:

Are there any concrete results you can show at this point?

* Yes, we have sample lyric generation and sample audios that we have ready. The results are not spectacular though.
  + Lyric generation:

Enter artist name (e.g., 'dua lipa'): coldplay

Enter genre (e.g., 'pop'): alternative

Enter optional starting text (or press Enter to skip):

Enter temperature (0.5-1.5, higher = more random): 1

Generating lyrics in the style of coldplay (alternative)...

==================================================

knee goofy karrueche boppers suffer woo weakness ahyea touchdown

pushin' spark richie brazilian sees accusations champion auburn empathize

blown liver men asks kahlúa dealers struggling drama filetofish

clean mo' heart closed wore wrestle identicalbut touch youre

masked engaged oohoohoohooh inner half swore trapped queen's because

ghosts signature digest giving about challengers spending dogs replayin'

live wealth insomnia htown tit wanna griffin fields rest

mekhi digest gone dyin' sleep arts map skinny freakazoid

brr em drankin' planes rainbows elevating disastrously stepsisters actin'

bringin' rapunzel groups ohohohohoh cheated tied whisper ugh noah

critiqued wassy children skip settling start alligators follow arab

rollin' texas preferably compromised trait kickin' burst under biggin'

jamaica wearin' quite headband stackin' provocative confronted hella shawty's

reminder stained two scent reat flavor tay whips ooouuu

bridesmaid ester skills blap rhyme messi puthy remote caged

beside dun sharpest blond dr yen rent's queens ears

artillery worst aziz fold trend color

* Audio generation results will be shown during the meeting.

How is your model performing compared with expectations?

* + For the transformer part, the implemented conditional transformer model demonstrates functional lyric generation capabilities despite quality limitations from the small training dataset. While the output shows random word combinations rather than coherent and stylistically accurate lyrics, the system successfully shows the core technical components such as the artists, genre, conditioning, and temperature by an interactive interface. The foundation is technically sound, with the primary path to improvement being dataset expansion and more sophisticated sampling strategies rather than architectural changes.
  + For the stable diffusion, the results are not favorable, but we do recognize that this could be attributed to the simple architecture of our model (not many hyperparameters), our limited gpu access, and our relatively small dataset. Effective image generation models, like stable diffusion, are very data hungry and are usually very large, so we did what we could given our limitations.

**Plan**: Are you on track with your project?

* + What do you need to dedicate more time to?
    - Improving our stable diffusion model results, as they seem too noisy at the moment.
    - Most of what is left is fine tuning.
  + What are you thinking of changing, if anything?
    - We are thinking about changing how noise is added as it seems to exceed the decibel ranges covered by spectrograms.
    - Minimal changes overall.