Extended Essay Outline

October

- Find what type of model I want to use.
- Figure out how what data

Interesting Articles:

https://openreview.net/pdf?id=BygXFkSYDH

Overall article explaining representation learning, current techniques and official definitions
https://neptune.ai/blog/representation-learning-with-autoencoder

Is very similar to my case. They use autoencoders to generate music embeddings. Some tasks they included were context prediction and composer classification from noisy data . The idea is that if the network is able to perform well in these tasks, they are able to extract meaningful music embeddings. They also used a piano-based dataset that had clips represented as piano rolls. https://arxiv.org/pdf/1706.04486.pdf

Uses convolutional autoencoders and 2-D spectrograms to generate music embeddings for each 2 seconds of the song: https://m-lin-dm.github.io/Deep_audio_embedding/

How to handle audio data by hugging face. There is even a course for this and could provide meaningful sources. If I want to use transformers in my project, I need an embedding network to create embeddings for each time sample. Each time sample is added with a positional encoding as well: https://huggingface.co/learn/audio-course/chapter1/audio_data

Talks about common techniques used in deep learning. Piano rolls are some easy way of representing music, but are however limited since it is difficult to show when a note ends or not: https://arxiv.org/pdf/1709.01620.pdf

https://link.springer.com/article/10.1007/s00521-019-04076-1

https://github.com/mlachmish/MusicGenreClassification

Some relevant datasets from international society of music:
https://www.ismir.net/resources/datasets/

This is the maestro dataset that contains lots 200+ hours of piano performances in .wav format
https://magenta.tensorflow.org/datasets/maestro

The following tensorflow teaches how to extract and plot the piano rolls of files from the maestro dataset; https://www.tensorflow.org/tutorials/audio/music_generation

The "next steps" section of the tensorflow tutorial led to me maybe using GANs for representation learning. This led me to this paper; https://arxiv.org/pdf/1612.08879.pdf

Turns out this idea of using GANs for representation learning has been explored in depth. In this book, it even talks about my initial idea of RNN autoencoders. They also discuss GANs. Perhaps I can find a pre-trained genre classification.
https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/ch17.html

I think I will be using MIDI datafiles for this project. It has note encodings and accounts for the spatial dimension of it. I found a lofi-playlist as well!
https://www.kaggle.com/datasets/zakarii/lofi-hip-hop-midi

https://arxiv.org/pdf/2106.05630.pdf MusicBERT: Talks about how an encoder-decoder structure. They formatted their input-features via piano rolls and other things that include pitch, time signature and duration.

Consider sentence embeddings as a relevant field. There exists some pre-trained models like the Universal Sentence Encoder: https://amitness.com/2020/06/universal-sentence-encoder/
https://medium.com/@busra.oguzoglu/sentence-embedding-methods-a-survey-7c62857f7b43

Changing thesis question to a comparison of different models. "Which input representation is better suited for generating music embeddings?"

Examines the effect of input data on the generated music embeddings. Ultimately, study should show how input representation affects a model's capability to generate audio embeddings which is representative of its understanding.

Paper that discusses different inputs and their effects on transformer learning. Although they used it for music generation, they visualized the chord embeddings and saw differences in their clusters.
https://arxiv.org/pdf/1606.04930.pdf

- Use same transformer model and parameters, but change each input representation in the following datatypes:
    - Image (mel spectrograms)
    - Waveform
    - Piano Rolls
- Two ways to analyze results: Accuracy in predicting music genre and visualization of embeddings.

Mel Spectrograms:

- Could use something similar to ViT transformers. The authors released different sizes for the model: small and large. I might use just 12 transformer encoder blocks since it is smaller, it's what the author's used and they based it off BERT Small: https://openreview.net/pdf?id=YicbFdNTTy
- Actual usage of ViT in audio spectrograms in Hugging face. I could maybe fine tune the model for 12 transformer encoding blocks since they also used the ViT transformer https://huggingface.co/docs/transformers/model_doc/audio-spectrogram-transformer

Waveform and Piano Rolls

- Use an architecture similar to BERT. 12 Encoding blocks and 12 self attention heads, since it has been used for music analysis before: https://archives.ismir.net/ismir2021/paper/000090.pdf (hugging face article) https://huggingface.co/blog/bert-101#3-bert-model-size--architecture
- https://huggingface.co/learn/audio-course/chapter4/fine-tuning#picking-a-pretrained-model-for-audio-classification

- Keras tutorials for audio classification. Can modify it to change it to music: https://keras.io/examples/audio/wav2vec2_audiocls/

If training from scratch could also analyze how long it takes the model to converge.

https://scholar.google.ca/scholar?hl=en&as_sdt=0%2C5&q=input+representation+in+music+machine+learning&btnG=
https://www.google.com/search?q=is+input+representation+important+in+audio+processing&sca_esv=575843313&rlz=1C1GCEB_enCA1043CA1043&sxsrf=AM9HkKnz-sSjUkjTIXb0bTPpN8U9eVH74w%3A1698083916677&ei=TLQ2Zan1KPLP0PEPl7WUiAk&ved=0ahUKEwipy5vF34yCAxXyJzQIHZcaBZEQ4dUDCBA&uact=5&oq=is+input+representation+important+in+audio+processing&gs_lp=Egxnd3Mtd2l6LXNlcnAiNWlzIGlucHV0IHJlcHJlc2VudGF0aW9uIGltcG9ydGFudCBpbiBhdWRpbyBwcm9jZXNzaW5nMgUQIRigAUiGQVDtBliEQHADeAGQAQGYAaoCoAHcJaoBBjUwLjQuMbgBA8gBAPgBAagCFMICBxAjGOoCGCfCAhYQABgDGI8BGOUCGOoCGLQCGIwD2AEBwgIWEC4YAxiPARjlAhjqAhi0AhiMA9gBAcICBxAjGIoFGCfCAgQQIxgnwgIIEAAYigUYkQLCAgsQABiABBixAxiDAcICERAuGIAEGLEDGIMBGMcBGNEDwgILEAAYigUYsQMYgwHCAgsQLhiKBRixAxiDAcICBxAAGIoFGEPCAgsQLhiABBixAxiDAcICCBAAGIAEGLEDwgIFEAAYgATCAggQABiABBjJA8ICCBAAGIoFGJIDwgIIEC4YgAQYsQPCAgoQABiABBhGGPkBwgIKEAAYgAQYFBiHAsICIRAAGIAEGEYY-QEYlwUYjAUY3QQYRhj0Axj1Axj2A9gBAsICCBAAGBYYHhgPwgIGEAAYFhgewgIIEAAYigUYhgPCAggQIRgWGB4YHcICBRAAGKIEwgIEECEYFcICChAhGBYYHhgPGB3CAgcQIRigARgKwgIEECEYCuIDBBgAIEGIBgG6BgYIARABGAu6BgYIAhABGBM&sclient=gws-wiz-serp

https://www.kaggle.com/datasets/jorgeruizdev/ludwig-music-dataset-moods-and-subgenres

https://www.justinsalamon.com/uploads/4/3/9/4/4394963/cramer_looklistenlearnmore_icassp_2019.pdf

URL list from Sunday, Oct. 15 2023 16:13 PM

To copy this list, type [Ctrl] A, then type [Ctrl] C.

Extended Essay Outline - Google Docs
https://docs.google.com/document/d/1GvWAUvJ_kOdBXe82voUN1b_swXhUhcZbCzo11Nfscc8/edit

transformer encoder-decoder - Google Search
https://www.google.com/search?q=transformer+encoder-decoder&oq=transformer+encoder-decoder&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTIHCAEQABiABDIICAIQABgWGB4yCAgDEAAYFhgeMggIBBAAGBYYHjIICAUQABgWGB4yCAgGEAAYFhgeMggIBxAAGBYYHjIICAgQABgWGB4yCAgJEAAYFhge0gEJMTA2NzNqMGo0qAIAsAIA&sourceid=chrome&ie=UTF-8

Genshin Impact 1 Hour Liyue Mix + Animated Background - YouTube
https://www.youtube.com/watch?v=s7gdfHquLzY&ab_channel=Fludix9

BERT in music classification - Google Search
https://www.google.com/search?q=BERT+in+music+classification&oq=BERT+in+music+classification&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTIHCAEQIRigATIHCAIQIRigAdIBCDg2MzFqMGo0qAIAsAIA&sourceid=chrome&ie=UTF-8

3_28-JIP.dvi
file:///Users/catherinebalajadia/Downloads/Melody2Vec_Distributed_Representations_of_Melodic_.pdf

universal sentence encoder architecture - Google Search
https://www.google.com/search?q=universal+sentence+encoder+architecture&sca_esv=572781667&tbm=isch&sxsrf=AM9HkKlekWFMRsheo31b38xCrvyMLktEsg:1697095530664&source=lnms&sa=X&ved=2ahUKEwiZ9ITC_e-BAxV6FTQIHcxoAN0Q_AUoAXoECAEQAw&biw=1440&bih=783&dpr=2#imgrc=ixrcHb-B7MEFRM

song representation learning - Google Search
https://www.google.com/search?q=song+representation+learning&sca_esv=572793137&sxsrf=AM9HkKlTTj-u0Zpl1T5baPRvPBdrsp99aQ%3A1697095896989&ei=2KAnZdT6O8iz0PEP8vuE6A0&ved=0ahUKEwjUxtvw_u-BAxXIGTQIHfI9Ad0Q4dUDCBA&uact=5&oq=song+representation+learning&gs_lp=Egxnd3Mtd2l6LXNlcnAiHHNvbmcgcmVwcmVzZW50YXRpb24gbGVhcm5pbmcyBRAhGKABMgUQIRigAUjSQ1DsBViLQnAKeACQAQCYAbgBoAGoH6oBBTEyLjI1uAEDyAEA-AEBqAIUwgIHECMY6gIYJ8ICFhAAGAMYjwEY5QIY6gIYtAIYjAPYAQHCAhYQLhgDGI8BGOUCGOoCGLQCGIwD2AEBwgIHECMYigUYJ8ICCBAAGIoFGJECwgIHEAAYigUYQ8ICERAuGIMBGMcBGLEDGNEDGIAEwgILEAAYgAQYsQMYgwHCAhEQLhiABBixAxiDARjHARjRA8ICBBAjGCfCAgUQABiABMICCxAuGIAEGLEDGIMBwgIHEC4YigUYQ8ICDRAAGIoFGLEDGIMBGEPCAg0QLhiKBRixAxiDARhDwgIOEAAYigUYsQMYgwEYkQLCAhAQABiABBgUGIcCGLEDGIMBwgILEC4YrwEYxwEYgATCAgsQLhiABBjHARjRA8ICBxAAGIAEGArCAgoQLhiABBjIBBgKwgINEC4YgAQYxwEYrwEYCsICDRAAGBYYHhgPGPEEGArCAggQABiKBRiGA8ICDRAuGIoFGMcBGNEDGEPCAggQLhiABBixA8ICExAuGIoFGLEDGIMBGMcBGNEDGEPCAgkQLhiKBRgKGEPCAg4QLhiABBixAxiDARjIBMICCBAAGIAEGLEDwgILEAAYigUYsQMYgwHCAg0QABiABBixAxiDARgKwgILEC4YgAQYxwEYrwHCAgUQLhiABMICChAAGIAEGBQYhwLCAgYQABgWGB7CAggQABgWGB4YD8ICFBAuGIAEGJcFGNwEGN4EGOAE2AECwgIKEAAYFhgeGA8YCsICCBAAGAgYHhgNwgIEECEYFcICBxAhGKABGAriAwQYACBBiAYBugYGCAEQARgLugYGCAIQARgU&sclient=gws-wiz-serp#ip=1

Pop2Piano
https://huggingface.co/docs/transformers/model_doc/pop2piano

BERT for music - Google Search
https://www.google.com/search?q=BERT+for+music+&sca_esv=572793137&sxsrf=AM9HkKnm7QJ2nD6N8jKN4jKokWM_BX4k0A%3A1697097014231&ei=NqUnZenaDY-T0PEP3ZG7-AI&ved=0ahUKEwjpzbqFg_CBAxWPCTQIHd3IDi8Q4dUDCBA&uact=5&oq=BERT+for+music+&gs_lp=Egxnd3Mtd2l6LXNlcnAiD0JFUlQgZm9yIG11c2IjIDIGEAAYFhgeMgYQABgWGB4yBhAAGBYYHjIKEAAYFhgeGA8YCjIIEAAYigUYhgMyCBAAGIoFGIYDSPwfUIIMWM8dcAJ4AZABAJgBlAGgAa8NqgEENS4xMbgBA8gBAPgBAagCFMICBxAjGOoCGCfCAhAQABiKBRjqAhi0AhhD2AEBwgIWEAAYAxiPARjlAhjqAhi0AhiMA9gBAsICBxAjGIoFGCfCAgQQIxgnwgIIEAAYigUYkQLCAgcQABiKBRhDwgIKEC4YigUY5QQYQ8ICCxAAGIAEGLEDGIMBwgIFEC4YgATCAgsQLhiDARixAxiKBcICBRAAGIAEwgIOEAAYigUYsQMYgwEYkQLCAgsQLhiDARixAxiABMICDRAuGIoFGLEDGIMBGEPCAgsQLhiABBjHARivAcICCxAuGK8BGMcBGIAEwgIIEAAYFhgeGA_CAggQABgWGB4YCuIDBBgAIEGIBgG6BgYIARABGAG6BgYIAhABGAs&sclient=gws-wiz-serp#ip=1

000090.pdf
https://archives.ismir.net/ismir2021/paper/000090.pdf

data2vec explained - Google Search
https://www.google.com/search?q=data2vec+explained&oq=data2vec+explained&gs_lcrp=EgZjaHJvbWUyBggAEEUYOdIBCDMyNDhqMGo0qAIAsAIA&sourceid=chrome&ie=UTF-8

GitHub - ruanchaves/song2vec: Telegram bot that recommends songs as YouTube playlists through gensim's word2vec
https://github.com/ruanchaves/song2vec

transfer learning genre classification - Google Search
https://www.google.com/search?q=transfer+learning+genre+classification&sca_esv=572793137&sxsrf=AM9HkKlALzmPjgt-AjUECeEHWCwypzPX-A%3A1697098018242&ei=IqknZZu1DsOL0PEP18C1qAw&ved=0ahUKEwibzprkhvCB

AxXDBTQIHVdgDcUQ4dUDCBA&uact=5&oq=transfer+learning+genre+classification&gs_lp=Egxnd3Mtd2l6LXNlcnAiJnRyYW5zZmVyIGxlYXJuaW5nIGdlbnJlIGNsYXNzaWZpY2F0aW9uMgYQABgIGB4yCBAAGIoFGIYDMggQABiKBRiGAzIIEAAYigUYhgNI8A1QjgpYywxwAngBkAEAmAGEAqABlAeqAQUwLjMuMrgBA8gBAPgBAcICChAAGEcY1gQYsAPiAwQYACBBiAYBkAYI&sclient=gws-wiz-serp

Music Genre Classification using Transfer Learning(Pytorch) | by Aryan Khatana | The Startup | Medium
https://medium.com/swlh/music-genre-classification-using-transfer-learning-pytorch-ea1c23e36eb8

nlp - Sum vs mean of word-embeddings for sentence similarity - Data Science Stack Exchange
https://datascience.stackexchange.com/questions/110718/sum-vs-mean-of-word-embeddings-for-sentence-similarity

arXiv:1405.4053v2 [cs.CL] 22 May 2014
https://arxiv.org/pdf/1405.4053v2.pdf

doc2vec gensim custon - Google Search
https://www.google.com/search?q=doc2vec+gensim+custon&oq=doc2vec+gensim+custon&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTIJCAEQIRgKGKABMgkIAhAhGAoYoAHSAQg0MzkxajBqN6gCALACAA&sourceid=chrome&ie=UTF-8

Doc2Vec Model — gensim
https://radimrehurek.com/gensim/auto_examples/tutorials/run_doc2vec_lee.html

A gentle introduction to Doc2Vec. TL;DR | by Gidi Shperber | Wisio | Medium
https://medium.com/wisio/a-gentle-introduction-to-doc2vec-db3e8c0cce5e