

LLM IN ARTS

Music Generation

Text-to-Music

By Abhigyan and Shri



Problem Statement

Creating an entire musical composition based on a given thematic prompt involves seamlessly blending musical elements and lyrical content to evoke specific emotions and themes.

This task requires a high level of creativity and expertise akin to professional music production in the entertainment industry.

The goal is to craft a cohesive and emotionally resonant song that captures the essence of the provided theme and musical style, showcasing a deep understanding of musical composition and lyrical storytelling.

Data Set Creation

- We used Chords Lyric Dataset available in Kaggle:
<https://www.kaggle.com/datasets/eitanbentora/chords-and-lyrics-dataset>
- It consist of chords, lyrics and genre.
- To create the prompts did reverse engineering.
- We used Pretrained Summarization model that is available on Hugging Face.
The model is: **knkarthick/bart-large-xsum-samsum**
- Later we cleaned the dataset to curate the chords.
- Our team would like to thank **Maria Pushparaj** from team **Pheonix** in helping us curating the Chords dataset and helping us in understanding them.

Project Phases

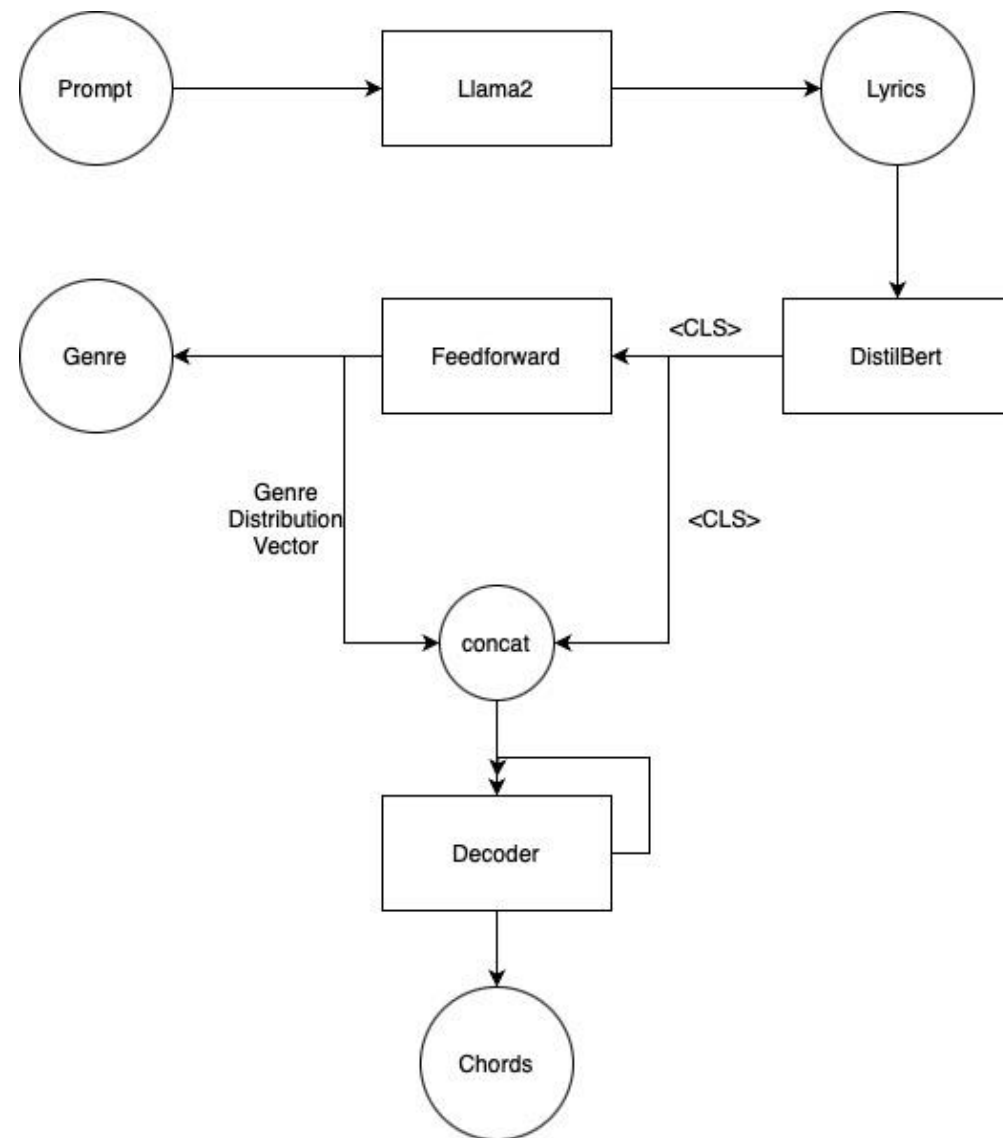
The project has 3 phases:

- Generating the Lyrics from the prompt.
- Classifying the Genre of the of the lyrics created.
- Generating the Chord Progression.



Project High Level Architecture

The system takes a prompt theme of the song and feeds it to a large language model (Llama2) to generate a Lyrics. This Lyrics is then used for two purposes: chord generation and genre distribution vector creation. DistilBert, classifies the lyrics to genre label. Finally, the refined genre and the lyrics pathway output are combined and fed to a decoder model that generates the Chord Progression.

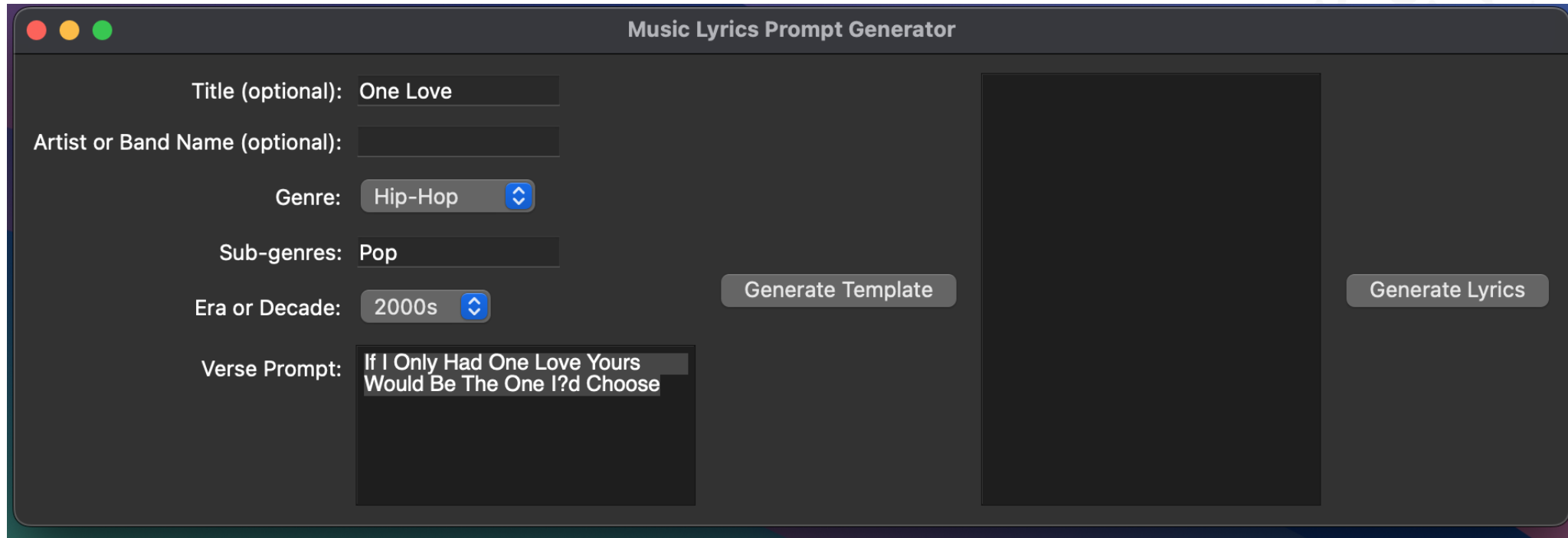


Phase-1: Lyrics Generation

- Experimented with FlanT5 and Llama 3 to create a finetuned model to generate lyrics.



Prompt generator



The screenshot shows a web application titled "Music Lyrics Prompt Generator" with a dark theme. It features several input fields and two buttons. The inputs are: "Title (optional):" with the value "One Love"; "Artist or Band Name (optional):" which is empty; "Genre:" with a dropdown menu showing "Hip-Hop"; "Sub-genres:" with the value "Pop"; "Era or Decade:" with a dropdown menu showing "2000s"; and "Verse Prompt:" with the text "If I Only Had One Love Yours Would Be The One I'd Choose". The "Generate Template" button is positioned between the input fields and a large empty text area. The "Generate Lyrics" button is located to the right of the large text area.

Music Lyrics Prompt Generator

Title (optional): One Love

Artist or Band Name (optional):

Genre: Hip-Hop

Sub-genres: Pop

Era or Decade: 2000s

Verse Prompt: If I Only Had One Love Yours
Would Be The One I'd Choose

Generate Template

Generate Lyrics

Prompt Generation

Title: [Provide a title or theme for the song]

Artist or Band Name: [Specify the name of the artist or band, real or fictional]

Genre: [Specify the primary musical genre or style]

Sub-genres: [List any relevant sub-genres or related styles]

Era or Decade: [Indicate the era or decade that influences the song's style]

Verse Prompt: [Provide a detailed prompt or key points for the first verse]

Additional instructions:

[Add any specific requirements or guidelines for the lyrics]

Example lyrics :

I Saw A Cat With A Machine In His Brain The Man Who Fed Him Said He Didn't Feel Any Pain I'd Like To See That Man Take Out That Machine And Stick It In His Own Brain You Know
What I Mean I Saw A Rabbit With Its Eyes Full Of Tears The Lab That Owned Her Had Been Doing It For Years Why Don't We Make Them Pay For Every Last Eye That Couldn't Cry Its
Own Tears Do You Know What I Mean When I Tell You That We'll All Be Looking For Changes Changes In The Way We Treat Our Fellow Creatures And We Will Learn How To Grow
Well I Tell You That We'll All Be Looking For Changes Changes In The Way We Treat Our Fellow Creatures And We Will Learn How To Grow When We're Looking For Changes I Saw A
Monkey That Was Learning To Choke A Guy Beside Him Gave Him Cigarettes To Smoke And Everytime That Monkey Started To Cough The Bastard Laughed His Head Off Do You
Know What I Mean When I Tell You That We'll All Be Looking For Changes Changes In The Way We Treat Our Fellow Creatures And We Will Learn How To Grow When We're Looking
For Changes We're Looking For Changes We're Looking For Changes We're Looking For Changes In The Way We Are

Find...

C

instruction "Write a song lyric based on the given inputs and verse prompt."

Output "lyrics We were inseparable Everything I had to do I did it next to you And the memories we made were so incredible Then our love was interrupted by my schedule There was nothing that I could do Cause you fell into the deepest depression baby And I hate to know I'm responsible And your heart filled up with so much aggression baby You got used to Being alone alone You adapted now your use to Being alone all alone Ooo you got used too being on your own I saw it happenin' But I didn't accept the truth I couldn't fathom it There was so much going on you couldn't handle it Could have divided my time I should have fractioned it But there was nothing that I could do Cause you fell into the deepest depression baby And I hate to know I'm responsible And your heart filled up with so much aggression baby You got used to Being alone alone (all alone) You adapted (you adapted) now your use to Being alone all alone (didn't mean to leave you lonely girl) Ooo you got used too being (used to) on your own (all on your own) t/t n"

instruction "Write a song lyric based on the given inputs and verse prompt."

```
input "genres:['canadian pop', 'pop', 'post-teen pop']progression:['Bm', 'E', 'F#m']start_keyDverse promptPrayed our love wasn't vain."
```

[illegible]

► 2

▶ 3

► 4

► 5

► 6

▶ 7

▶ 8

9

▶ 10

11

12

▶ 13

Phase-2: Genre Classification

We used pretrained BERT based models to classify the genre.

Experiments and Metrics:

Model	Loss	Accuracy
bert-base-uncased	1.27	0.62
distillbert-base-uncased	1.16	0.74
veucci-lyrics-to-genre	1.38	0.49
twitter-roberta-base-sentiment-latest	1.32	0.58
j-hartmann/emotion-english-distilroberta-base	1.36	0.52

Phase-2: Genre Classification

We went with DistilBERT as it is simple model and giving better accuracy.

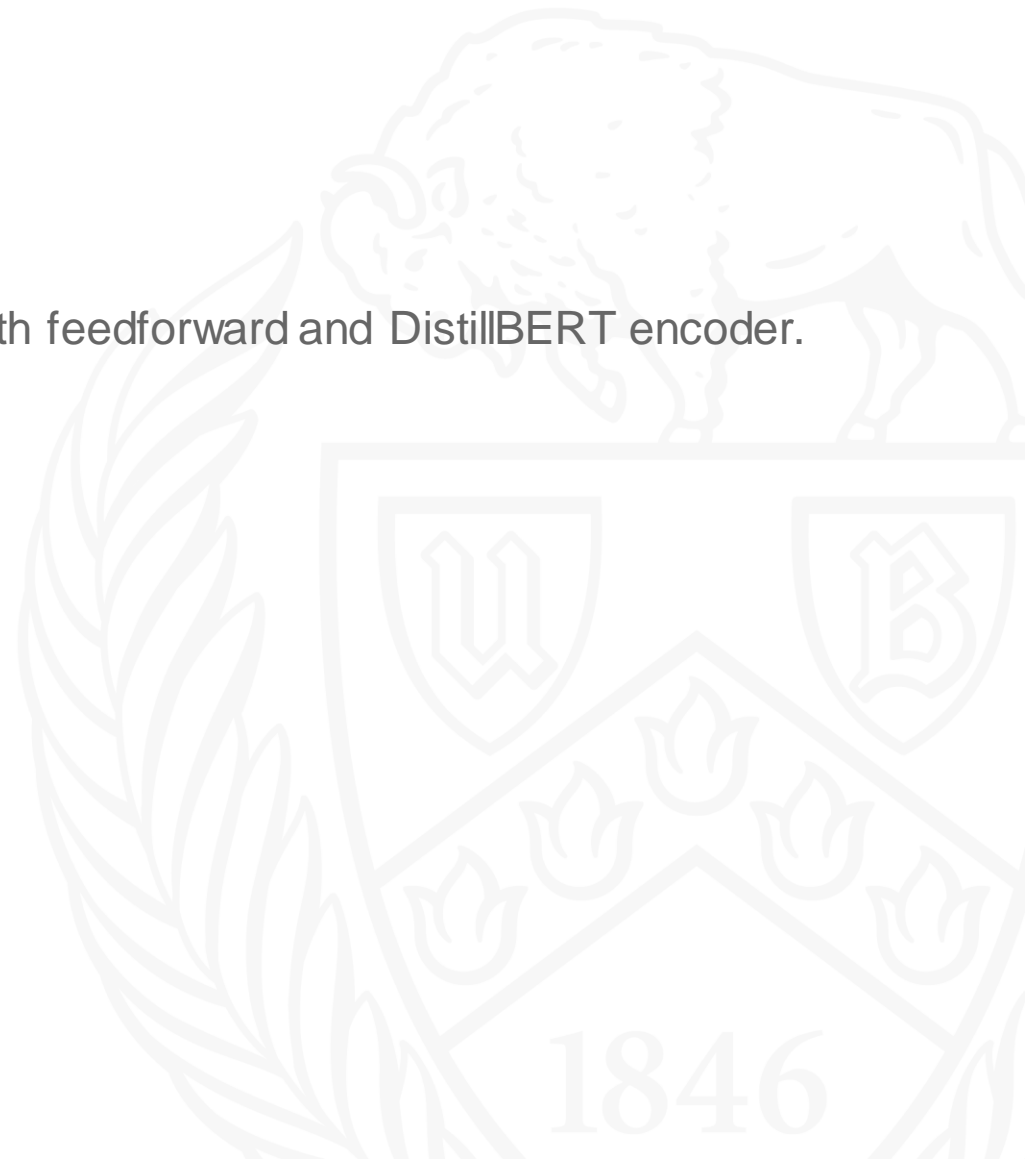
Examples:

Song Name	Predicted Genre	Original Genre
Vertical Horizon	Rap	Rap
Godflesh	Metal	Metal
Word Wizard	Country	Pop
8th of November	Pop	Country

Phase-2: Genre Classification

Improvements:

- Instead of finetuning only the feed-forward classifier we can fine both feedforward and DistillBERT encoder.
- This will result in increase in Accuracy.
- We can collect more data which is diverse.



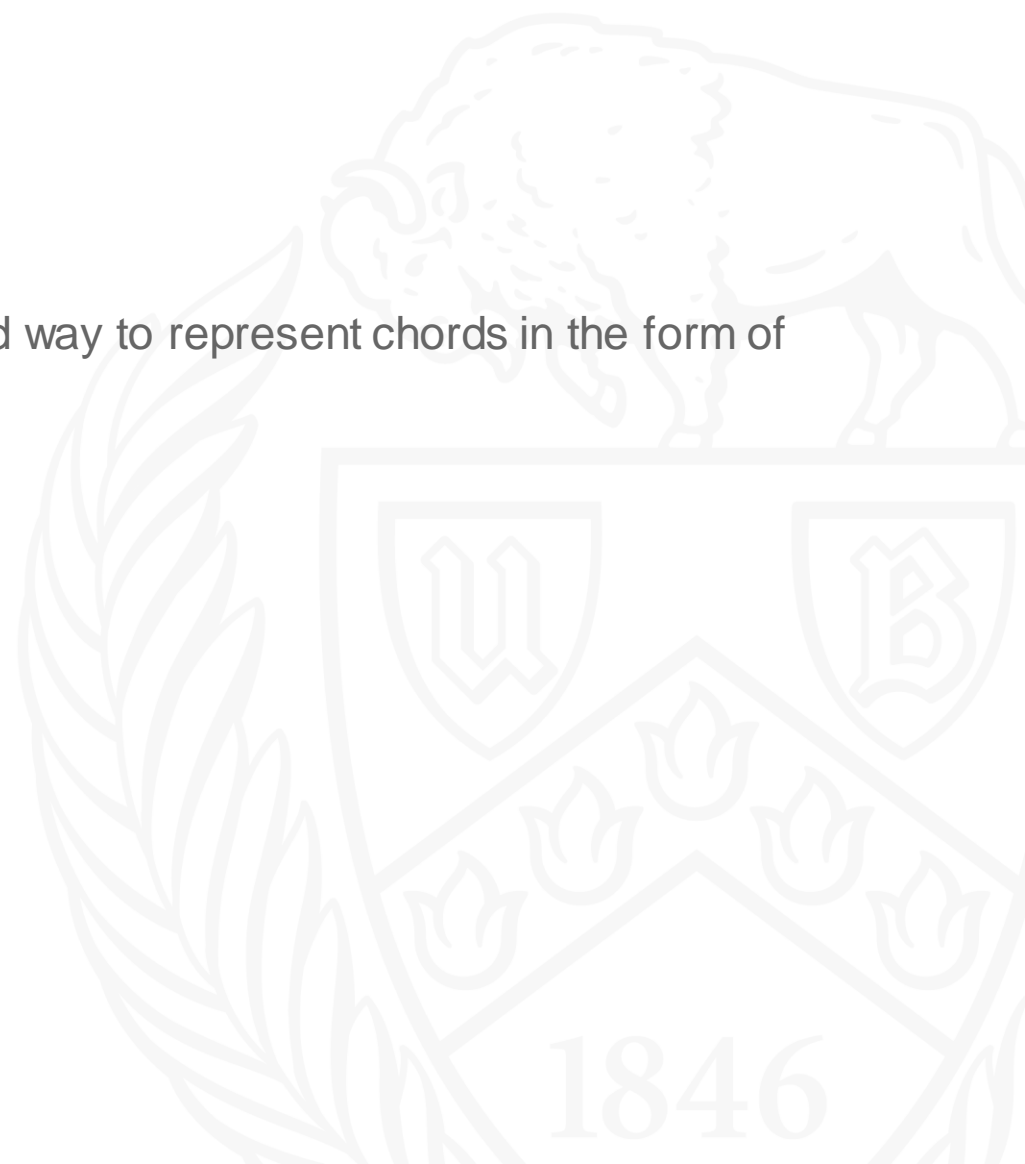
Phase-3: Chord Progression

Challenges:

- Chord Representation. We don't have any direct way or pre-defined way to represent chords in the form of embeddings.
- So, using Transformers to extract chords is out of the equation.
- We will use One-Hot encoders to represent chords.

To generate Chord Progression, we had two approaches:

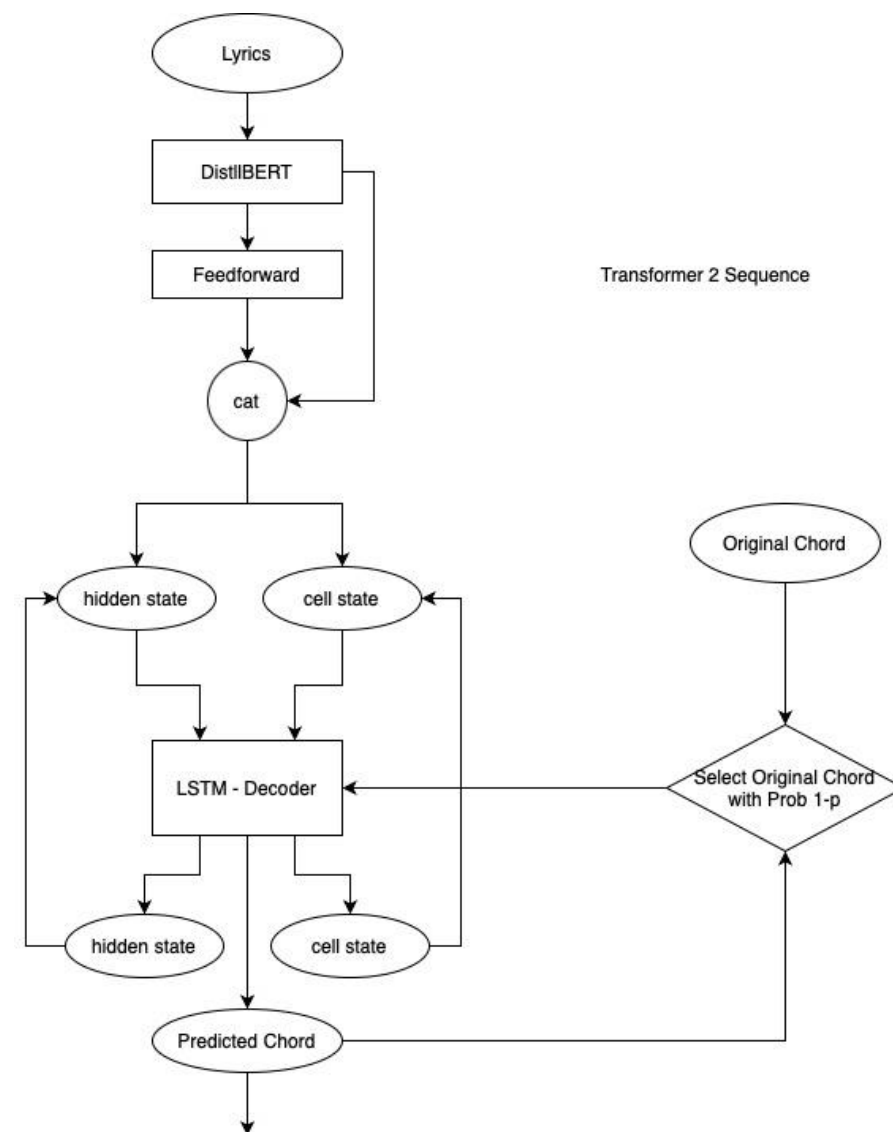
- Transformer-2-Sequence Encoder-to-Decoder approach.
- RHLF using Policy Gradient



Phase-3: Chord Progression

Transformer-2-Sequence Encoder-to-Decoder:

- Used Encoder and Decoder Paradigm
- Transformer used in Genre is freezed Encoder.
- LSTM is decoder.
- We train only the Decoder.
- Used Teacher Force Ratio to speed up the convergence.



Phase-3: Chord Progression

Transformer-2-Sequence Encoder-to-Decoder:

Examples:

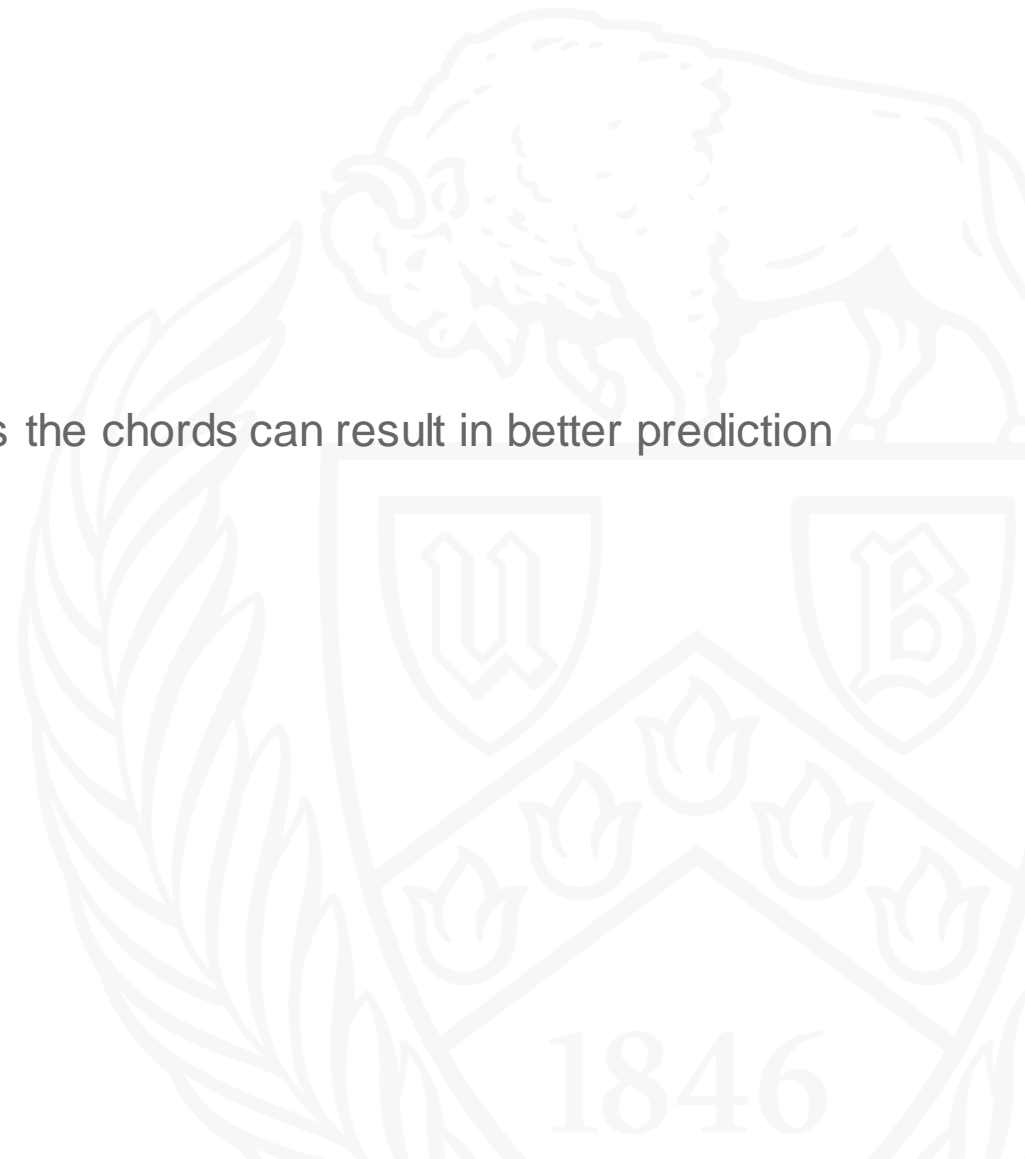
Song Name	Start Key	Predicted Progression	Original Progression
Who I have always been	Cadd9	Bb, A9, B/A	G, A, D
Don't let me be misunderstood	Am	Gbdim7, A/G#, Am/D	Dm, Am, Dm
Different World	Em	F9, Dsus, Fm/C	C, D, Bm
Where did my baby go	G	A7/G, B/C, G#m7	Em7, G, Em7

Phase-3: Chord Progression

Transformer-2-Sequence Encoder-to-Decoder:

Improvements:

- Along with Genre and Lyrics introducing other features that explains the chords can result in better prediction of the Chords.



Phase-3: Chord Progression

RHLF using Policy Gradient:

It has 2 steps:

- Construction Reward Model.
- Construct the Policy to select the chord given the state(Lyrics and Chords used at previous step).



Phase-3: Chord Progression

RHLF using Policy Gradient:

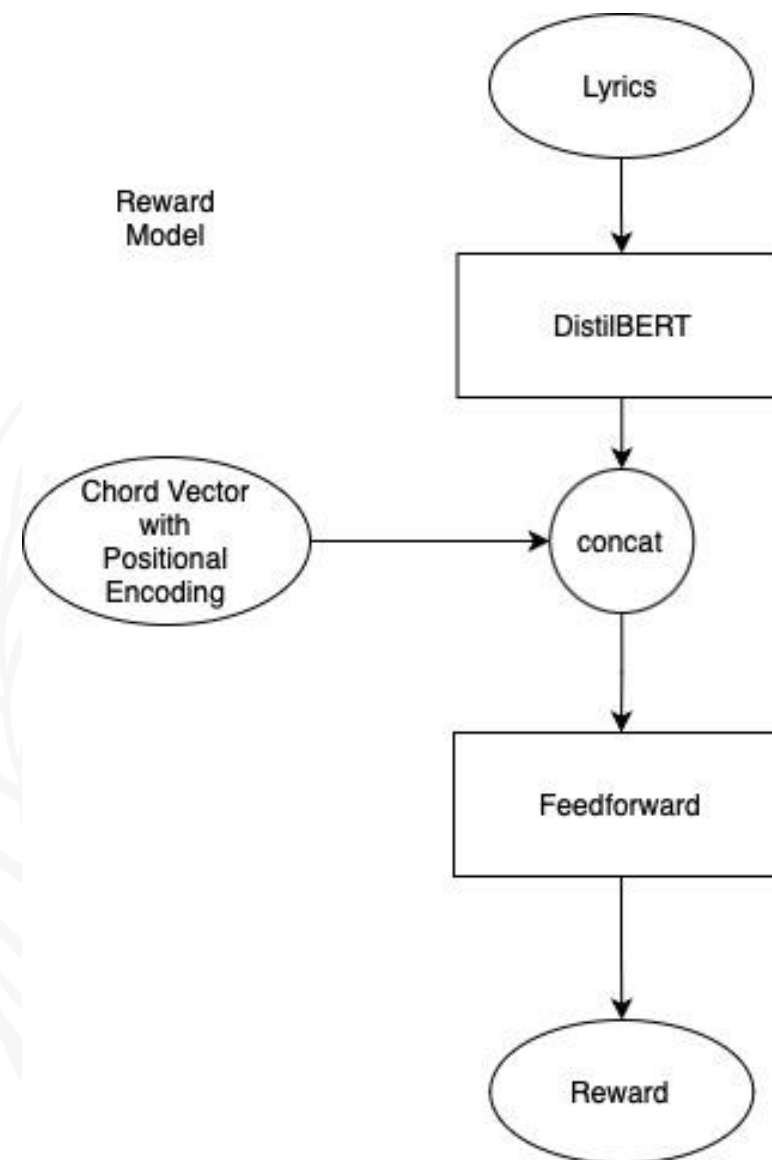
Construction of Reward Model:

Challenges:

- Representation of Chords in a single vector using Positional Encoding.

We overcame it adding 0 to first chord, 0.25 to second chord, 0.5 to third chord and 0.75 to fourth chord to one hot vector. So, the value at a particular chord will be 1, 1.25, 1.5 and 1.75.

- Curate Reward Data with Human Feedback.



Phase-3: Chord Progression

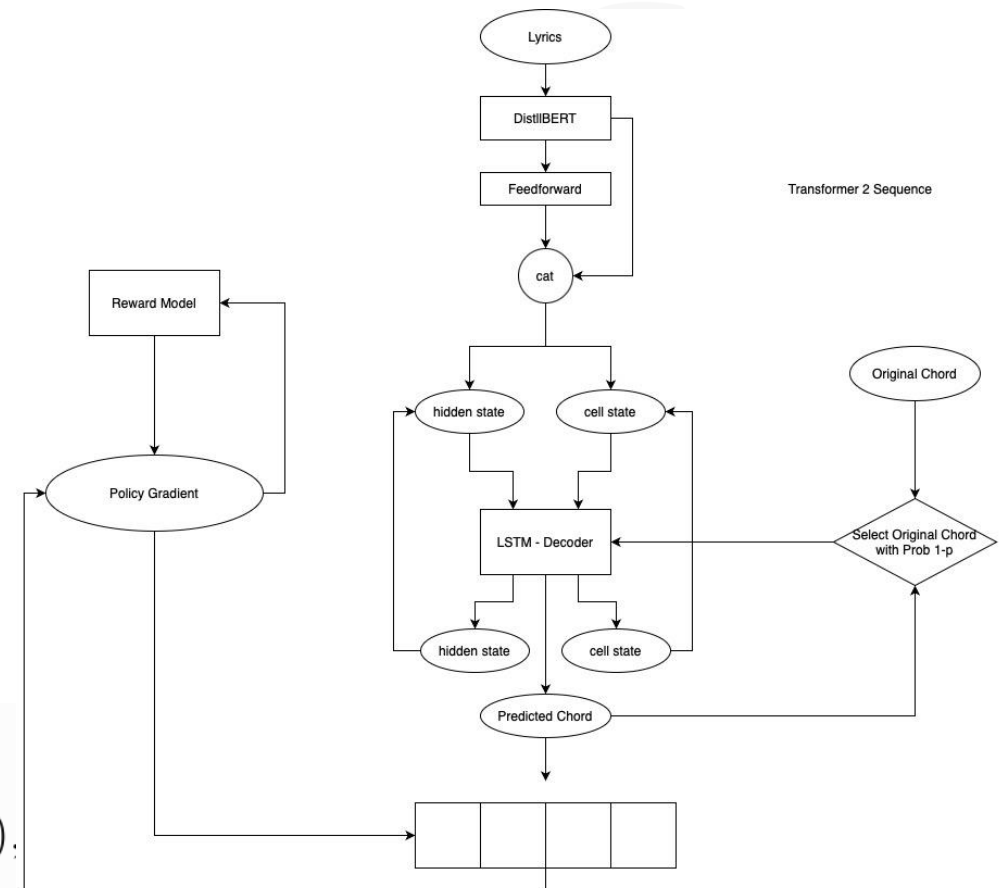
RHLF using Policy Gradient:

Construction of Policy:

- Using Policy Gradient to construct policy.
- Using the Reward generated by the reward model as the feedback is used to quantify the goodness of the action.

Loss Function: Maximize

$$\hat{g} = \frac{1}{|\mathcal{D}|} \sum_{\tau \in \mathcal{D}} \sum_{t=0}^T \nabla_{\theta} \log \pi_{\theta}(a_t | s_t) R(\tau),$$



Phase-3: Chord Progression

RHLF using Policy Gradient:

Examples:

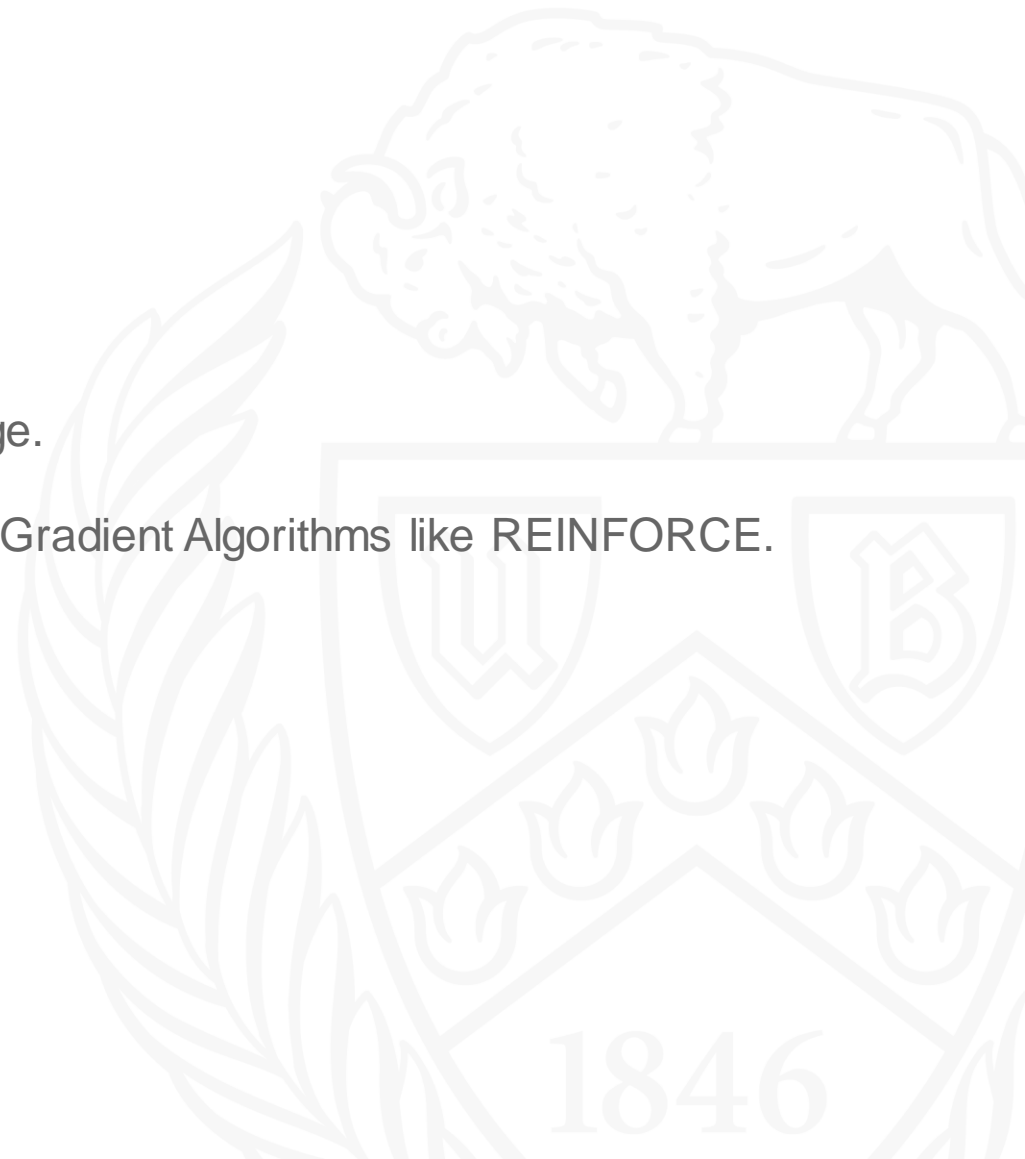
Song Name	Start Key	Predicted Progression	Original Progression
Ronan	C	Em, Am, Am7/E	Am, G, F
The Neighborhood	Em	E11, D/E, Dsus	G, C, Am
Wild Horses	G	D7, E4, A7m	G7, C7, Dsus9
So, will I	F	F#m/B, Cm, Absus2	Gm/F, C/F, Bb/F

Phase-3: Chord Progression

RHLF using Policy Gradient:

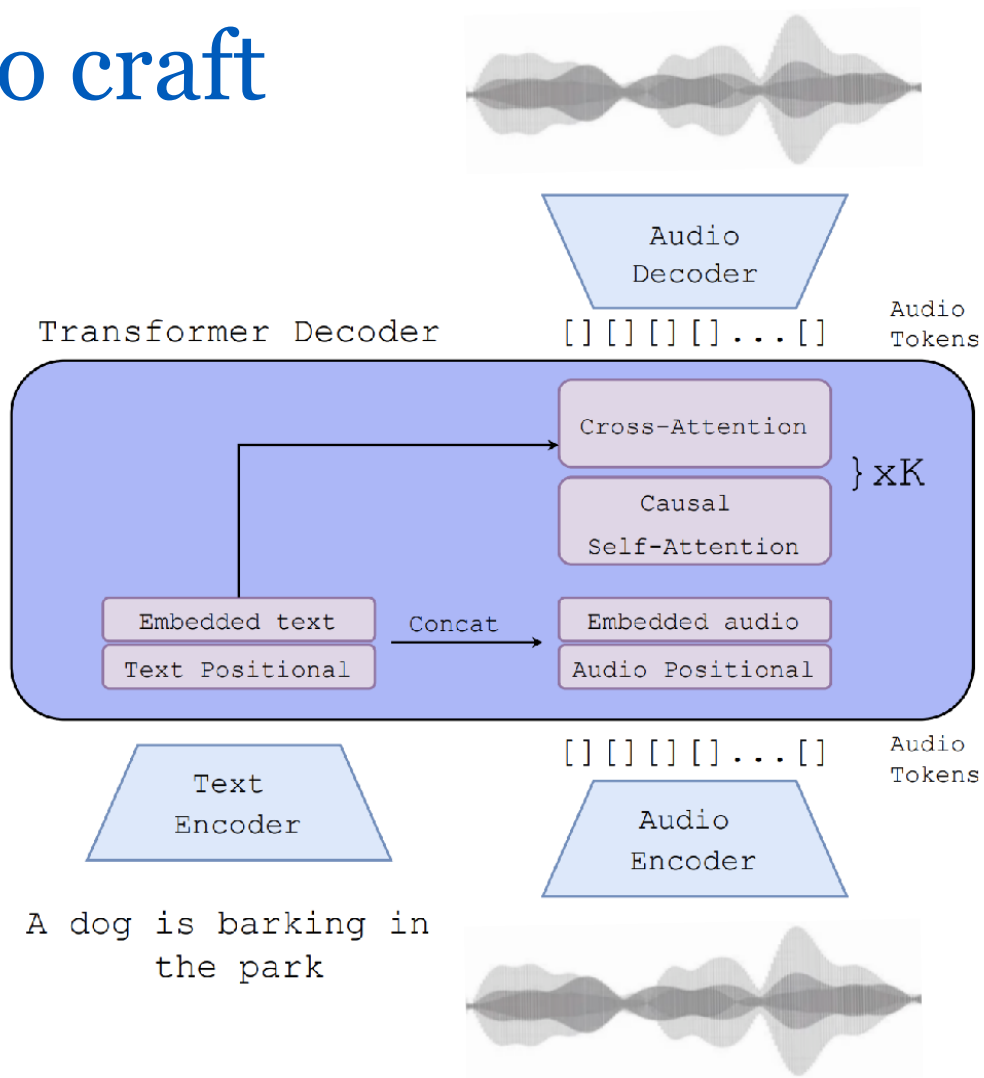
Improvements:

- Better Human Reward Construction using Chords domain knowledge.
- Introducing Intermediate Rewards so that we can use better Policy Gradient Algorithms like REINFORCE.

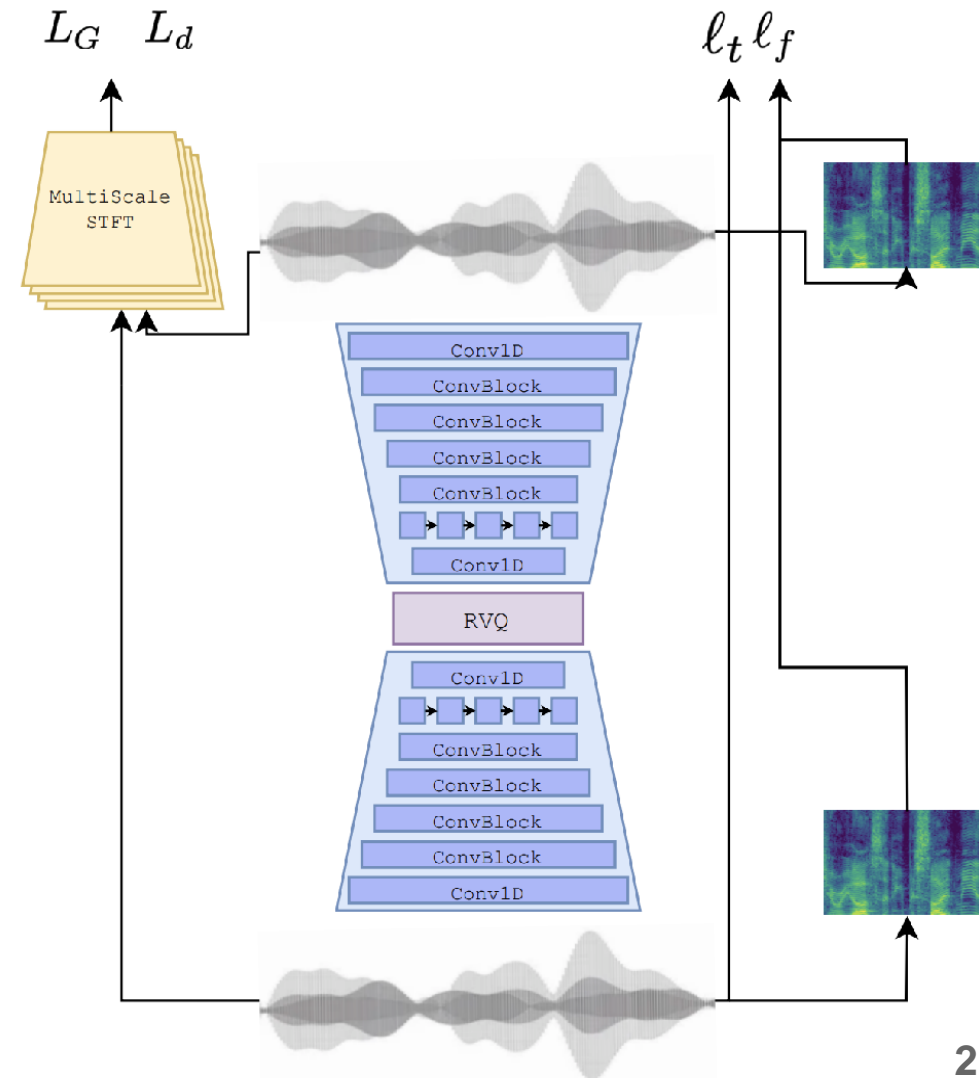


Audio craft

AudioGen Audio Representation Model



AudioGen Language Model






Audio Craft

- Audio gen only for small model (3B+)
- Prompt based music samples generated
 - Ex: Ar Rahman music styled tabla
- Can also use synthetic functions and text prompts together to create samples.
- Ex:
 - `torch.cos(2 * math.pi * 440 * t)`
 - Prompt:
['Jazz jazz and only jazz', 'Hindi tabla with flute and sitar',]



Audio prompting

- Audio as seed to generate music
- Multi Band Diffusion approach implemented in the neural encoder to generate closer music to the given sample
 - Generated music
 - Generated diffused music

Melody conditional generation

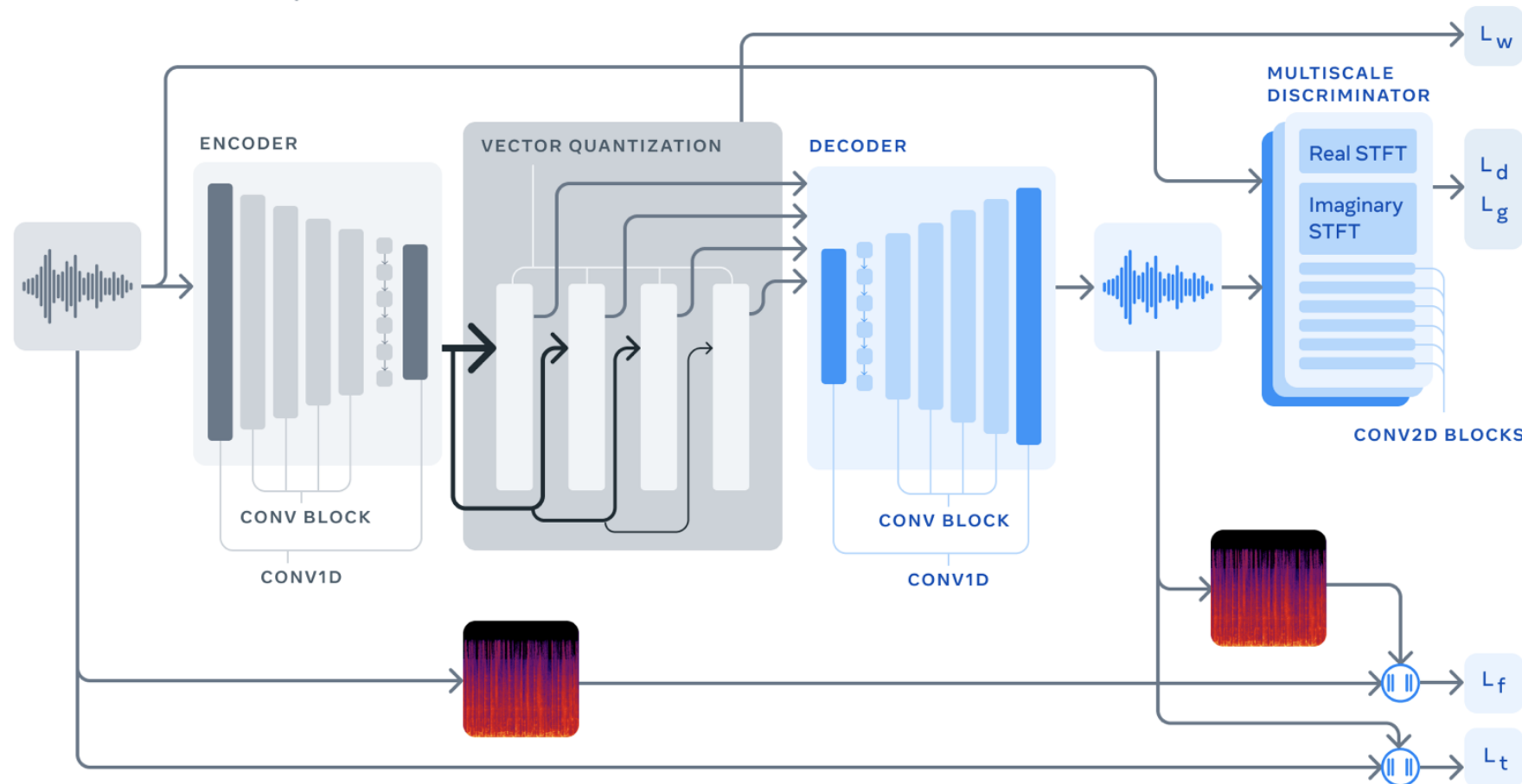
- AR Rahman , conditoned on
 - '90s bollywood music with deep instrumentals and base, hindi lyrics and very existential meaning',
 - 'Hindi tabla with flute and sitar'
- Sunset House Music mix in three phases , intro, mid and end



Genre clustering in embedding space

Meta AI's Audio Compression Process

- In progress
- Clustering in embedding space through the fb neural encoder model.



Song composer

Lyric to Melody

Generate a melody that matches the given lyrics.

“Speaking words of wisdom let it be”

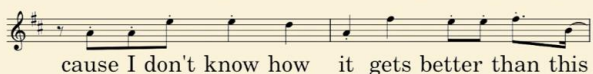
Melody to Lyric

Write lyrics that harmonize with this melody.



Song Continuation

Continue the song based on the given prompt.

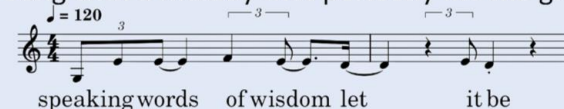


Text to Song

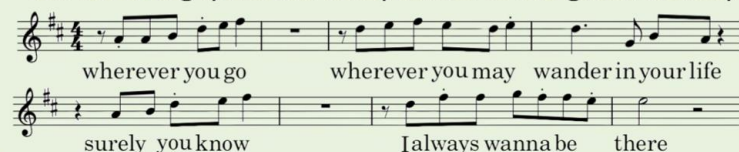
Create a song on brave and sacrificing with a rapid pace.



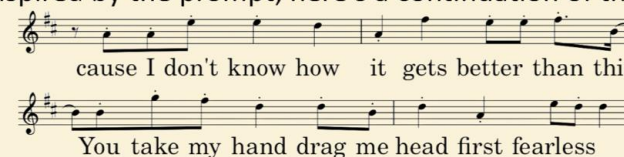
Here's a song with a melody that perfectly fits the given lyrics.



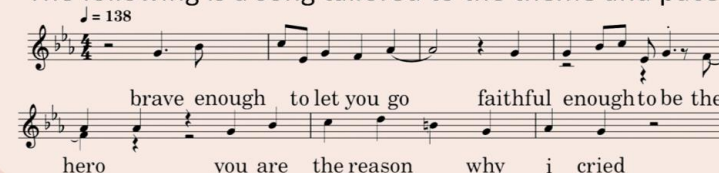
The following lyrics are finely tuned to the given melody.



Inspired by the prompt, here's a continuation of the song.



The following is a song tailored to the theme and pace.



Demo

- Lyrics to Melody:
- I care how you get here, just get here if you can.
You can reach me by sailboat,
climb a tree and swing a rope,
or cross the border in a blaze of hope.
I care how you get here.



Exciting Demo

- Prompt:

I want to create a song for my natural language class, taught by Professor Rohini Srihari and Teaching assistant Sayantal Pal.
Create lyrics for such a song , include the course name into the lyrics and have a line saying that they want to grade us very highly

- Lyrics Created

- NLP, the class that sets us free
Prof. Srihari and Sayantal, our guiding team

Natural Language Processing, oh so fine
They want to grade us high, all the time

From words to meaning, we learn it all
In this class, we stand tall

Parsing, disambiguating, we're on a roll
NLP, you're in our soul

Natural Language Processing, oh so fine
They want to grade us high, all the time

Prof. Srihari, Sayantal Pal, we thank you
NLP, our passion true



Questions

- Thank you

