RhymeCraftAI:
AI for lyrics generation

Data 255 - Deep Learning
Guided by:
Prof. Masum, Mohammad

Presented by:
Bhavik Patel
Abdul Sohail Ahmed
Smeet Sheth
Poojan Gagrani

Executive Summary

- **Project Objective**: The goal of the project is to create a lyric-generating application by utilizing the Genius Lyrics dataset. The application uses deep learning architectures, like Transformer models and LSTM networks, and sophisticated natural language processing (NLP) techniques to generate original lyrics that accurately convey the desired emotional and thematic nuances.

- **Innovation and Utility**: The project aims to investigate the relationship between creativity and technology, going beyond technical advancements, and offering a tool that improves musical composition and expression. It allows the music enthusiasts and composers, enabling customization based on mood, and artist influence.

- **Technical Approach**: Several deep learning models, beginning with simple GRUs and progressing to more intricate structures like Bi-LSTM and Transformer models with attention mechanisms, are implemented and evaluated in a stepwise manner throughout the project. These models were selected based on how well they handle sequential data and can pick up on the subtle syntactic differences found in lyrics.

- **Performance Evaluation**: The effectiveness of the models is assessed using metrics such as Rouge and BLEU scores to evaluate linguistic quality and relevance of generated lyrics.

- **Impact and Future Scope**: The project aims to demonstrate how AI can be utilized in artistic domains, contributing to computational creativity. With feedback systems and continuous learning, the application is expected to improve over time, providing insights into how machines can augment human creativity in music.

Project Background

Motivation

- The problem of creative block or lack of fresh inspiration
- Lacks of songs suitable for a particular theme
- Lyrics Generation process can be time-consuming

Needs and Importance

- Innovation in Music Production
- Enhancement of artistic expression
- Listeners want frequent new songs

Targeted Problem

- Assisting artists in lyrics generation
- Capturing the artist's style and theme
- Balancing Creativity and Coherence

Deliverables

- Final Report
- Final Presentation Slides
- Source Code

Project Requirements

Data Requirements	AI/ML Requirements	Functional Requirements
Access to large amount of data that contains details like artist name, genre, title, lyrics, language etc. Data should be accurately categorized	Different models like GRUs, BiLSTM and Transformer (GPT-2) should be employed to effectively handle different aspects of the lyrics and generate contextually relevant and sound lyrics	Users should be prompted for the artist name, theme and starting part of the lyrics (seed text)
Data should have the lyrics of different artist generated in such a way that it helps to provide information about its style	Proper training of the models needed to be performed so that they can capture the semantic information and lyrics complexities effectively	Users must be able to select the artist from the list and the themes should be provided related to that artist only
Data should be in proper format to be processed and prepared for the purpose of the modeling	Thematic information identified through topic modeling perfomed using Latent Dirichlet Allocation (LDA) should be properly integrated with lyrics generation process	Seed text should be the part of the lyrics and the generated lyrics should have only 100 words without any gibberish characters

Literature Survey

Author	Title	Dataset Used	Model(s) Used	Key Evaluation Results
Q. Xiao, H. Liu, H. Luo, M. Wang, Y. Shen	Lyrics & Tune Style Transfer	Songs taken from Lyric.com (3,865 lines of lyrics from 86 songs of original artist)(5,540 lines of lyrics from 77 songs of target artist)	Encoder and Decoder: RNN + GRU Cells Classifier: BiLSTM	Baseline Model (BiLSTM): BLEU score: 0.713 BERTscore : 0.918
A. Dirik, H. Donmez, P. Yanardag	Controlled cue generation for play scripts	1511 English play scripts	GPT 2-FT, PPLM+LDA, PPLM+CueDisc, PPLM+Emotion	GPT2+FT - Unigram Sim: 0.42, Bigram: 0.29 PPLM+CueDisc- Unigram Sim: 0.72, Bigram: 0.60
J. Domala, M. Dogra, A. Srinivasaraghavan	Lyrics Inducer using Bidirectional Long Short-Term Memory Networks	Songs scraped from the web (approx 13,000 lines)	LSTM, GRU, BiLSTM	Training Accuracy: LSTM - 0.815 GRU - 0.8114 BiLSTM - 0.826
K. Jesse	Implementation of Sequence to Sequence For Lyric Generation	Lyrics from metrolyrics.com	GRU (Sequence to Sequence)	BLEU score: 0.248

Project Resources

Hardware Requirements		
Hardware	**Memory**	**Usage**
12 Core CPU, 19 Core GPU, 16 Core Neutral Engine	RAM: 16 GB SSD: 1TB	Data handling, training, and lyrics generation
16 Core GPU	RAM: 128 GB SSD: 1TB	Model Training
T4 GPU	RAM: 16 GB	Model Training

Software Requirements		
Environment/Packages/Libraries	**Usage**	**Version**
Python	Code script for data cleaning, transformation, topic modeling and modeling	3.10.4
Tranformers	Offers pre-trained transformer models	4.35.2
WordCloud	Generate word cloud visualization	1.9.2
TensorFlow	Used for building and training DL models	2.15.0
Keras	Used for building and training DL models	3.0.0
Pandas	For data manipulation	2.1.1
NumPy	For numerical calculation	1.19.2
NLTK	Used to evaluate generated lyrics	3.7
Matplotlib	To create visualizations	3.8.0

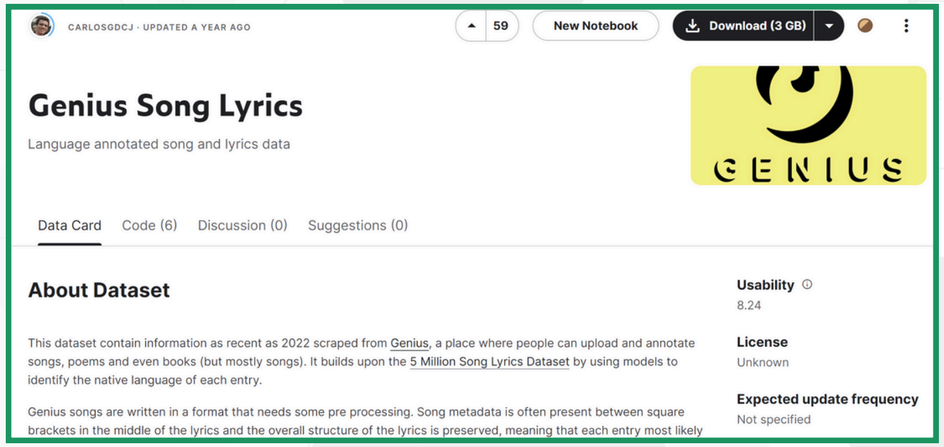
Project Resources

Tools and Licenses		
Tools	**Usage**	**Licenses**
Google Collab	To perform model building, training and lyrics generation	Proprietary
Jupyter Notebook	For initial data cleaning, transformation and Topic modeling	BSD-3 Clause
Excel	CSV files	Proprietary
Google Suite	For Project Documentation	Proprietary
Discord	To conduct virtual meetings	Proprietary
Grammarly Premium	To check if documents are grammatically correct or not	Proprietary
Google Drive	To store project resources	Proprietary

Resources and Costing			
Resources	**Justification**	**Duration**	**Costing**
Discord	For virtual meeting	4 months	Free
Grammarly	To check for any grammatical errors	4 months	$120
Google Collab	For modeling	3 months	$30

Data Sources

Genius Song Lyrics Dataset:



Feature overview

Column	Meaning
title	Title of the piece. Most entries are songs, but there are also some books, poems and even some other stuff
tag	Genre of the piece. Most non-music pieces are "misc", but not all. Some songs are also labeled as "misc"
artist	Person or group the piece is attributed to
year	Release year
views	Number of page views
features	Other artists that contributed
lyrics	Lyrics
id	Genius identifier
language_cld3	Lyrics language according to CLD3. Not reliable results are NaN
language_ft	Lyrics language according to FastText's langid. Values with low confidence (<0.5) are NaN
language	Combines language_cld3 and language_ft. Only has a non NaN entry if they both "agree"

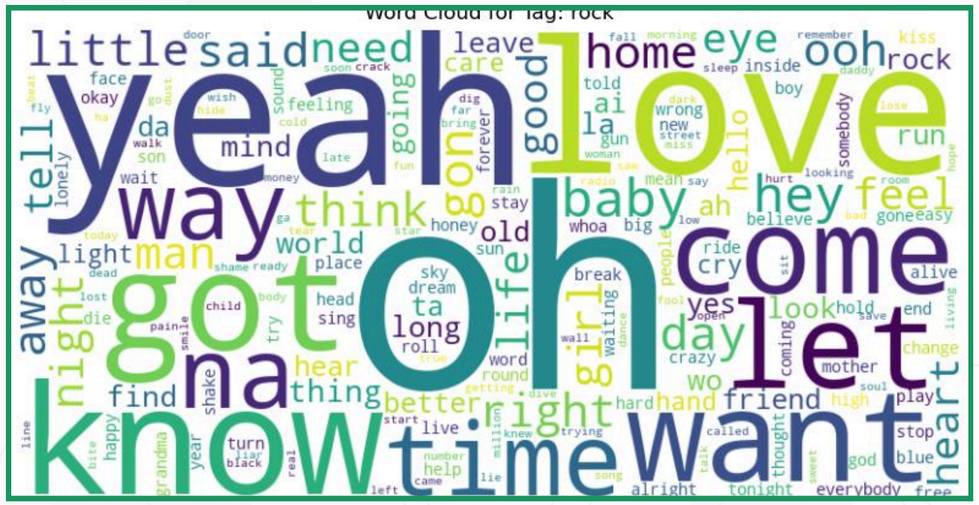
Data Statistics

No. of instances Before cleaning: 3,316,116

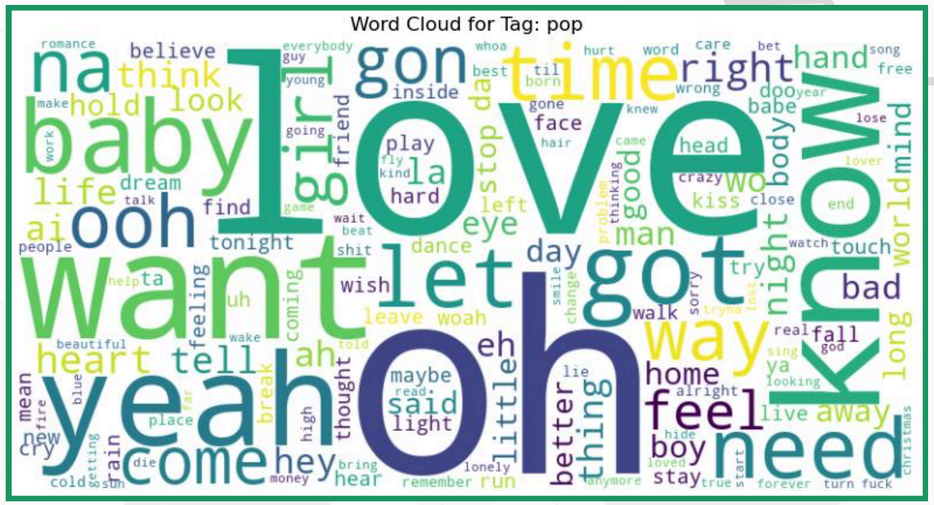
	Number of Unique Artists	Number of Unique of Songs	Number of Unique Tags	Null Values	Number of Unique Features
Raw Data	429,044	3,316,116	6	0	11
Clean Data	22	6,391	3	0	3 (Artist, Processed Lyrics, Theme)

EDA

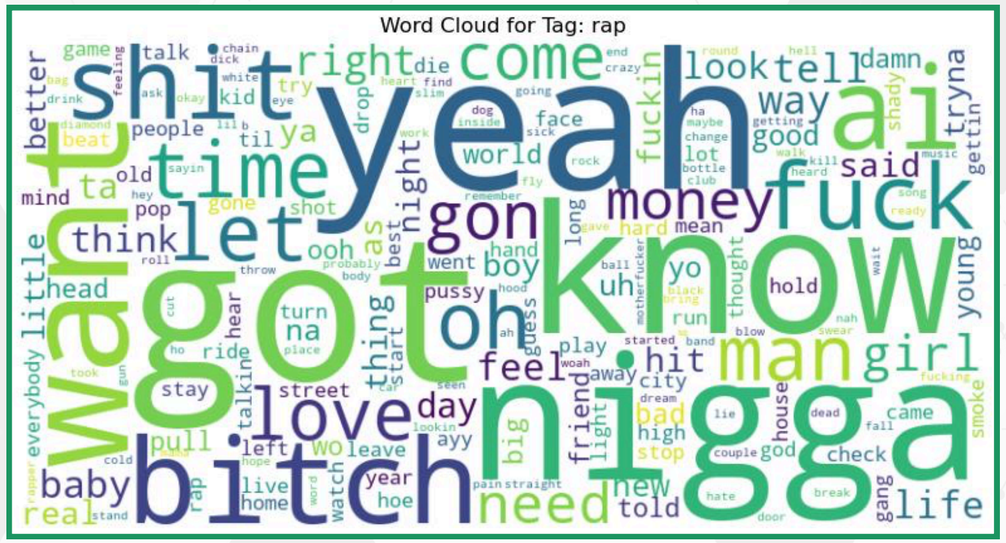
Rock



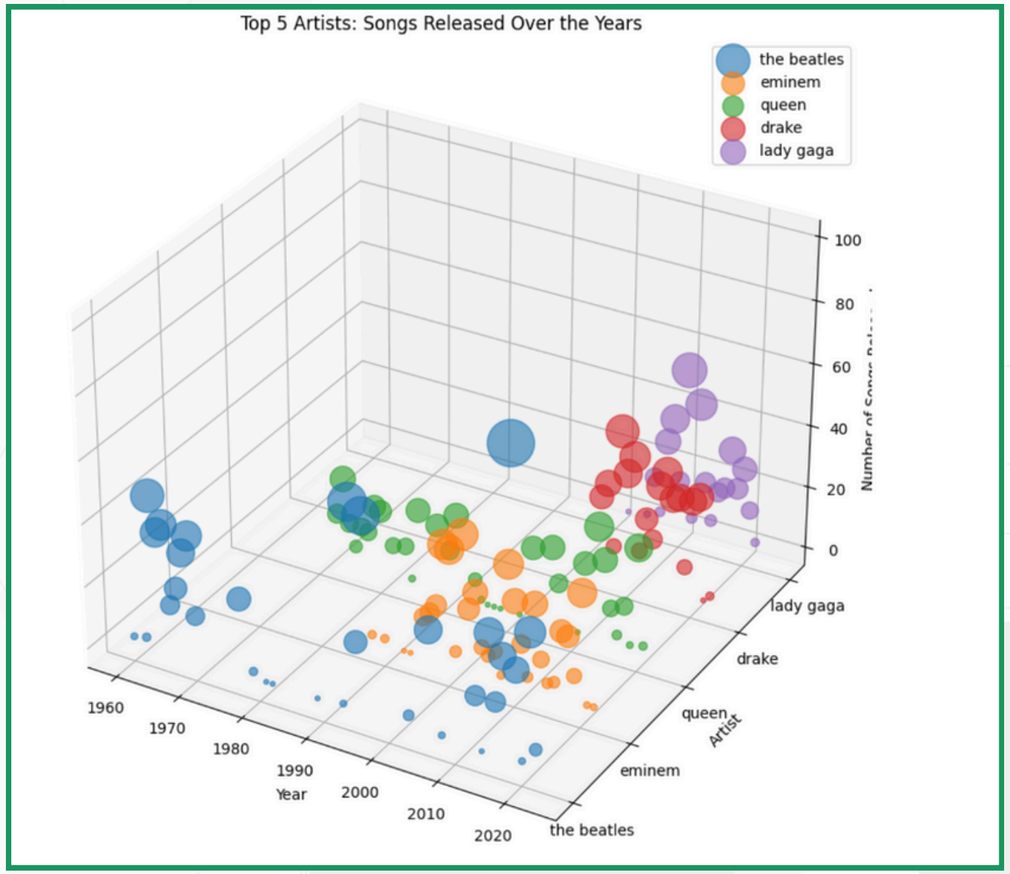
Pop



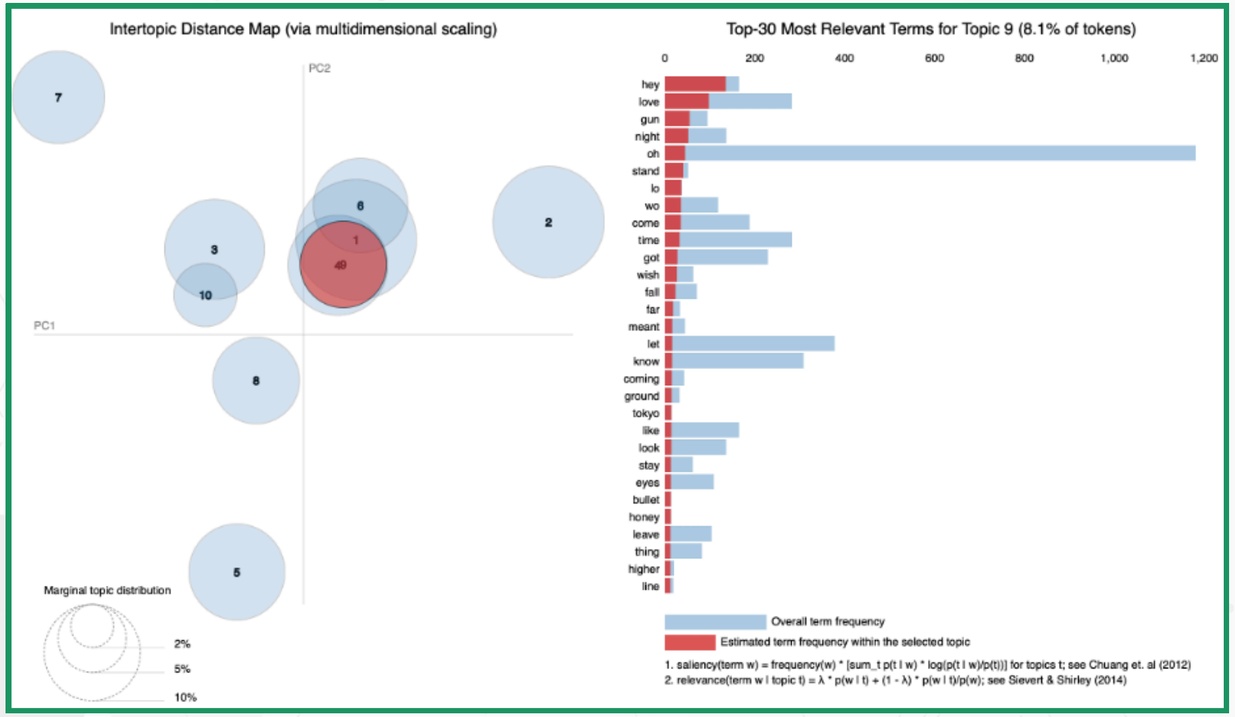
Rap



EDA



Songs released by top 5 Artists over the years.



Frequency of words in each topic compared to overall frequency.

Data Preprocessing

Genius Lyrics Dataset

Data Processing

1) Dropping Null Values
2) Dropping unnecessary features ('features', 'id', 'language_cld3', 'language_ft')
3) Adding space between punctuation, removing extra white space and data In brackets [...]as well as asterisk *...*

Data Subsetting

1) Keeping songs in only English Language
2) Sub setting data and keeping songs in tags ['pop', 'rap', 'rock']
3) Selecting top 22 artists
4) Selecting song release between year 1960 and 2024

LDA

1) Data preparation for LDA using NLTK, removing special characters, whitespaces, stop words etc.

2) Generating tokens, corpus and preparing topics using LDA

3) Generating theme for topics

Modeling

Project Workflow

Call "user_interaction()"

```
Available artists:
21 savage
artic monkeys
ariana grande
billie eilish
cardi b
dj khaled
drake
ed sheeran
eminem
Enter an artist from the list: halsey
```

Load Artist File

Select LDA theme

```
Available themes:
Intense Emotion
Bold and Rebellious
Reflection and Depth
Select a theme from the list: Intense Emotion
```

Sample tokenization

```
Original text: i'm searching for something that
Tokenized text: [7, 616, 24, 205, 14, 1, 58, 534
```

Representation of Embedding (2 x 25)

```
Embedding for 'i'm searching for something that':
0.109 0.141 -0.003 0.312 -0.181 0.104 0.311 -0.057 0.158 -(
0.132 0.162 0.257 0.149 0.085 0.034 0.336 -0.175 -0.114 0.1
-0.169 0.020 -0.108 -0.140 0.137 0.227 -0.332 -0.014 0.253
-0.168 -0.247 -0.097 0.056 -0.036 1.093 0.275 -0.024 0.121
```

Training Model

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 30, 50)	71300
bidirectional (Bidirectional)	(None, 30, 200)	120800
dropout (Dropout)	(None, 30, 200)	0
bidirectional_1 (Bidirectional)	(None, 200)	240800
dense (Dense)	(None, 100)	20100
dense_1 (Dense)	(None, 1426)	144026

Total params: 597026 (2.28 MB)
Trainable params: 597026 (2.28 MB)
Non-trainable params: 0 (0.00 Byte)

Give Seed

```
Model trained successfully. Please enter a seed phrase to generate lyrics.
Enter a seed phrase of up to 5 words to start the lyrics: one love
```

Generated song by RhymeCraftAI for Halsey Fans

```
One love you know i dated says they can't be of i know the sound i
know the sound of the boys wanna dancin' on top it yeah the boys
stop callin' the one sleeping in i'll on the stars and i keep knows
about to about you about you i still mean i know what i came where
the sound that you'll should know the sound i know the sea you love
you back mornin' i can't believe my man that i was make a brain and
i won't ever wrap of me not i hate you say the sound of
```

Evaluation
BLEU, ROUGE

Model 1 : glove.twitter.27B.100d + Bidirectional-GRU

Model Architecture

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 50, 100)	281500
bidirectional_5 (Bidirectional)	(None, 50, 64)	25728
batch_normalization_5 (BatchNormalization)	(None, 50, 64)	256
dropout_5 (Dropout)	(None, 50, 64)	0
bidirectional_6 (Bidirectional)	(None, 50, 64)	18816
batch_normalization_6 (BatchNormalization)	(None, 50, 64)	256
dropout_6 (Dropout)	(None, 50, 64)	0
bidirectional_7 (Bidirectional)	(None, 50, 64)	18816
batch_normalization_7 (BatchNormalization)	(None, 50, 64)	256
dropout_7 (Dropout)	(None, 50, 64)	0
bidirectional_8 (Bidirectional)	(None, 50, 64)	18816
batch_normalization_8 (BatchNormalization)	(None, 50, 64)	256
dropout_8 (Dropout)	(None, 50, 64)	0
bidirectional_9 (Bidirectional)	(None, 64)	18816
batch_normalization_9 (BatchNormalization)	(None, 64)	256
dropout_9 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 64)	4160
dense_3 (Dense)	(None, 2815)	182975

Hyperparameter Table

Hyperparameter Name	Value
Dropout rate	0.3
Number of GRU layers	5
Kernal Regualarizer	L2(0.001)
Dense layer Activation	ReLU
Output layer Activation	Softmax
Loss function	'categorical_crossentropy'
Optimizer	Adam with learning rate 0.002
Accuracy metric	'accuracy'
Number of epochs	200
Batch size	128

Model 2 : Sentence BERT + BiLSTM

Model Architecture

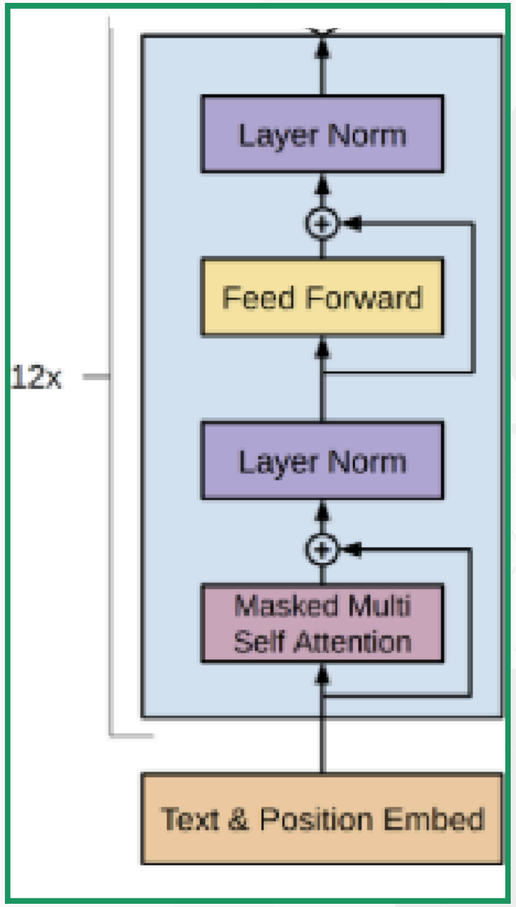
Layer (type)	Output Shape	Param #
`embedding` (Embedding)	(None, 30, 50)	71,300
`bidirectional` (Bidirectional)	(None, 30, 200)	120,800
`dropout` (Dropout)	(None, 30, 200)	0
`bidirectional_1` (Bidirectional)	(None, 200)	240,800
`dense` (Dense)	(None, 100)	20,100
`dense_1` (Dense)	(None, 1426)	144,026

Hyperparameter Table

Hyperparameter Name	Value
Activation function (Hidden)	ReLU
Activation function (Output)	Softmax
Weight Initializer	Default (Glorot Uniform)
Number of Hidden Layers	3 (2 LSTM layers, 1 Dense layer)
Neurons in Hidden Layers	100 in LSTM layers, 100 in Dense layer
Loss Function	Categorical Crossentropy
Optimizer	Adam
Number of Epochs	100
Batch Size	128
Learning Rate	Default of the Adam optimizer (0.001)
Evaluation Metric	Accuracy

Model 3 : Transformer based GPT2

Model Architecture



Layer (type)	Output Shape	Param #
wte (Token Embeddings)	(50257, 768)	38,597,376
wpe (Position Embeddings)	(1024, 768)	786,432
drop (Dropout)	-	-
h (Transformer Block x 12)	-	-
├─ ln_1 (Layer Normalization)	(768,)	768
├─ attn (Multi-Head Attention)	-	-
│ ├─ c_attn (Conv1D for QKV)	-	1,769,472
│ ├─ c_proj (Conv1D for output)	-	589,824
│ ├─ attn_dropout (Dropout)	-	-
│ └─ resid_dropout (Dropout)	-	-
├─ ln_2 (Layer Normalization)	(768,)	768
└─ mlp (Feedforward Network)	-	-
├─ c_fc (Conv1D to expand)	-	2,359,296
├─ c_proj (Conv1D to reduce)	-	2,359,296
├─ act (GELU Activation)	-	-
└─ dropout (Dropout)	-	-
ln_f (Layer Normalization final)	(768,)	1,536
lm_head (Linear)	(50257,)	38,597,376

Hyperparameter Table

Hyperparameter Name	Value
Activation Function (Hidden Layer)	GeLU
Activation Function (Output Layer)	Linear
Weight Initializer	Normal Distribution (std=0.02)
Number of Hidden Layers	12
Neurons in Hidden Layers	768
Loss Function	Categorical Cross-Entropy
Optimizer	AdamW
Number of Epochs	3
Batch Size	32
Learning Rate	5e-5
Evaluation Metric	Bleu, Rouge
Dropout Rate	0.1

Comparison of The Models

	glove.twitter.27B.100d + Bi-GRU	Sentence BERT + BiLSTM	GPT 2
Description	This model leverages the strengths of embedding , BiLSTM and gated recurrent unit architectures to process sequences of data.	Combines Sentence BERT used for capturing the sentence embedding with Bi-directional LSTM which is used for generating sequence of lyrics	Powerful autoregressive transformer model used for predicting next word in the sequence or lyrics.
Advantages	Integrating GloVe embeddings with Bi-GRU utilizes pre-trained semantic knowledge from a large Twitter dataset.	1. Strong at maintaining the thematic consistency of lyrics. 2. Effectively handle dependencies present in the sequence.	1. Capable of generating creative and complex lyrics 2. Can be tuned to properly capture the artist styles and lyrics themes.
Disadvantages	1. Might introduce biases present in the Twitter corpus 2. Complex and requires high computational power	1. Complex setup and requires proper tuning. 2. Highly resource intensive. 3. Struggle in generating highly creative lyrics.	1. Requires extensive amount of resources. 2. Possibility of generating themes is often biased without proper tuning.
Training time	2137.31 sec (T4 GPU)	189.68 sec (T4 GPU)	No Training
Inference time	9.56 sec (T4 GPU)	6.75 sec (T4 GPU)	4.02 sec (T4 GPU)
Evaluation	BLEU Score: 2.373 × 10-168 ROUGE: ROUGE-1: [‘Recall’ : 0.0679 ‘Precision’ : 0.6134, ‘F1' : 0.1222]	BLEU: 3.815 × 10-2 ROUGE: ROUGE-1 [‘Recall’ : 0.1935 ‘Precision’: 0.3429 ‘F1 Score’ : 0.2474]	BLEU: 0.1516 ROUGE: ROUGE-1 [‘Recall’: 0.1948 ‘Precision’: 0.375 ‘F1 score’: 0.2553]

Conclusion and Future work

Conclusion

- Development of melodic lyrics that match the user preferences
- Making the life of the artist less stressful by generating lyrics preserving specific themes and artist styles
- Successful development of deep learning models that generate appropriate lyrics

Future Work

- Provide 66 saved model for faster generation (21 artist x 3 themes)
- Enhancement of generated lyrics using more NLP techniques
- Providing the functionality of more customization by allowing users to generate more complex lyrics by combining multiple criteria
- Bringing in the concept of audio lyrics by using transformer models like Bark
- Creating more cultural friendly and diverse lyrics to suit the needs of larger audience

References

[1] Dirik, A., Donmez, H., & Yanardag, P. (2021). Controlled Cue Generation for Play Scripts. https://doi.org/10.48550/arXiv.2112.06953

[2] Domala, J., Dogra, M., & Srinivasaraghavan, A. (2021). Lyrics Inducer Using Bidirectional Long Short-Term Memory Networks. *In Proceedings of the 2021 International Conference on Communication and Computational Technologies,* 11-21. http://dx.doi.org/10.1007/978-981-16-3246-4_2

[3] Jesse, K. (n.d.). Implementation of Sequence to Sequence for Lyric Generation. https://kevinjesse.github.io/Resources/Lyrics/LyricNMT.pdf

[4] Xiao, Q., Liu, H., Luo, H., Wang, M., & Shen, Y. (n.d.). Lyrics & Tune Style Transfer. https://www.honghuluo.com/images/projects/song-style-transfer.pdf

Thank you