LYRICS BY GENRE : Classification and Generation using BERT and LSTM

NLP FINAL PROJECT PRESENTATION

Chloe Circenis, Gavin Hanville, Ima Mervin, Mia Ray, Maiana Vadas-Arendt

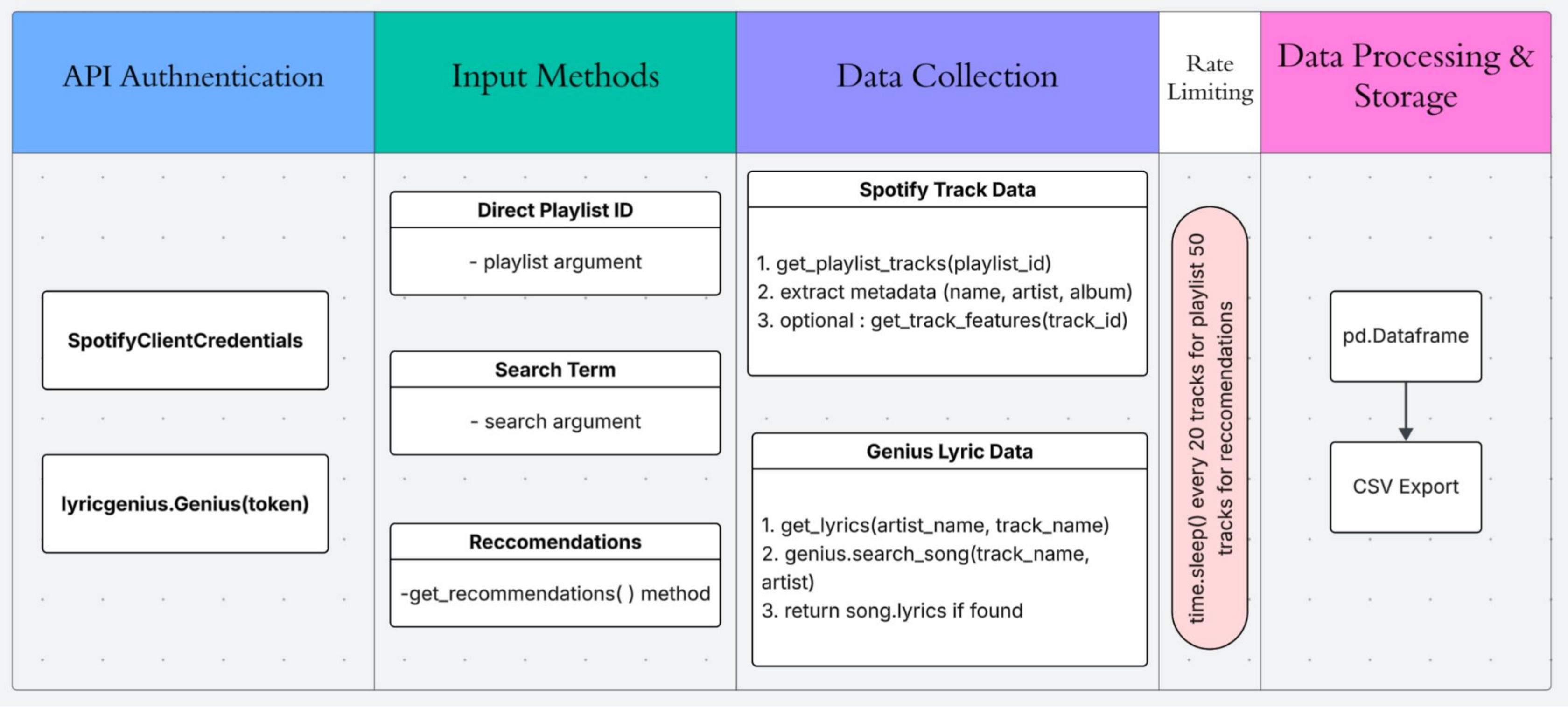
PROBLEM STATEMENT

How can we automatically classify and generate song lyrics by genre using NLP techniques?

MOTIVATION

Music is a culturally rich medium, and genre classification has applications in recommendation systems, music analysis, and creative AI. Generating genre-consistent lyrics is a challenging yet meaningful task in creative AI research

DATA COLLECTION



TASK FORMULATION

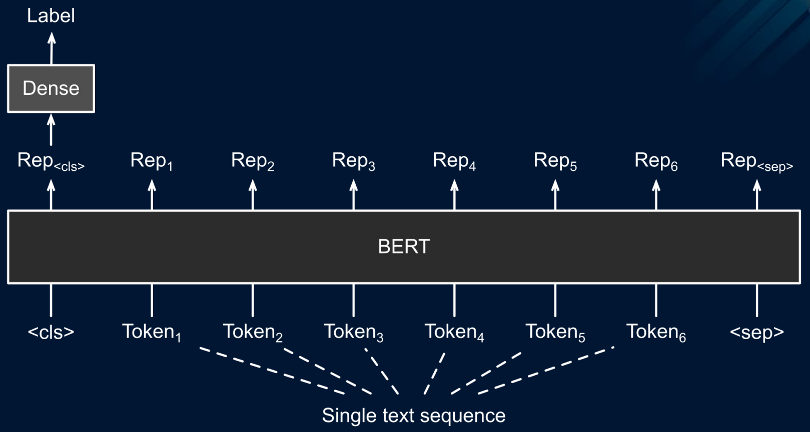
Using BERT for Sequence Classification

Both models are trained using song lyrics and the song's corresponding genre

Inputs are song lyrics

- ❖ Block Text as a string
- ❖ BERT Tokenizer (30,500 Tokens)

Outputs are the predicted song genre



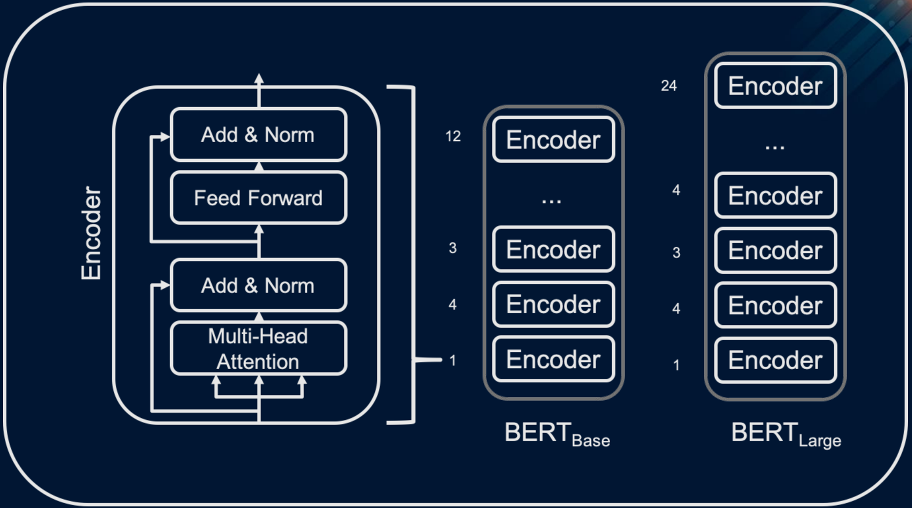
https://en.wikipedia.org/wiki/BERT_(language_model)

MODEL SELECTION

BERT
- ❖ Uncased
- ❖ Base
- ❖ For Sequence Classification

Provides Bi-Directional Context

Baseline Model Performance:
- ❖ Accuracy:
- ❖ Precision:
- ❖ F-1:



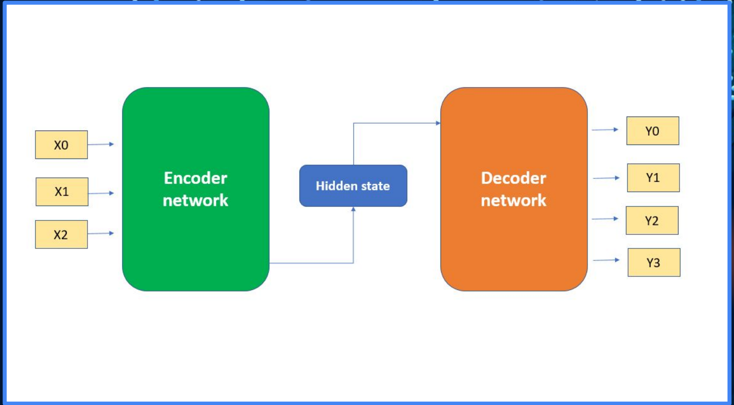
https://humboldt-wi.github.io/blog/research/information_systems_1920/bert_blog_post/

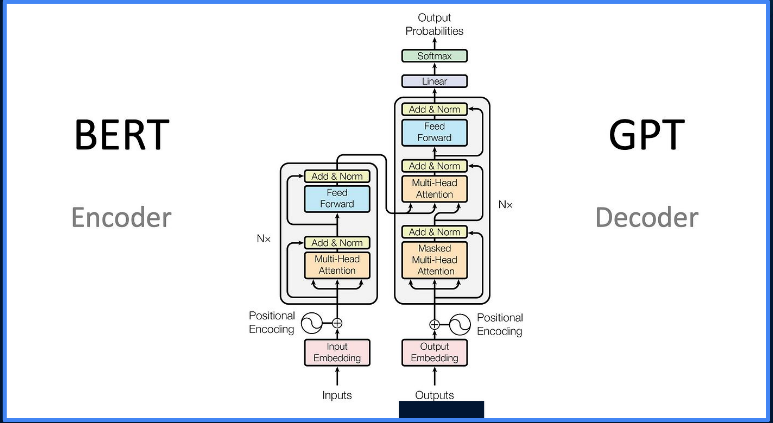
MODEL SELECTION

Encoder-Decoder

What and Why Encoder-Decoder?

- Originally popularized for **machine translation**
- Well suited for generating structured sequences – **LYRICS!!**





TRAINING DETAILS

Consistently Updating Library

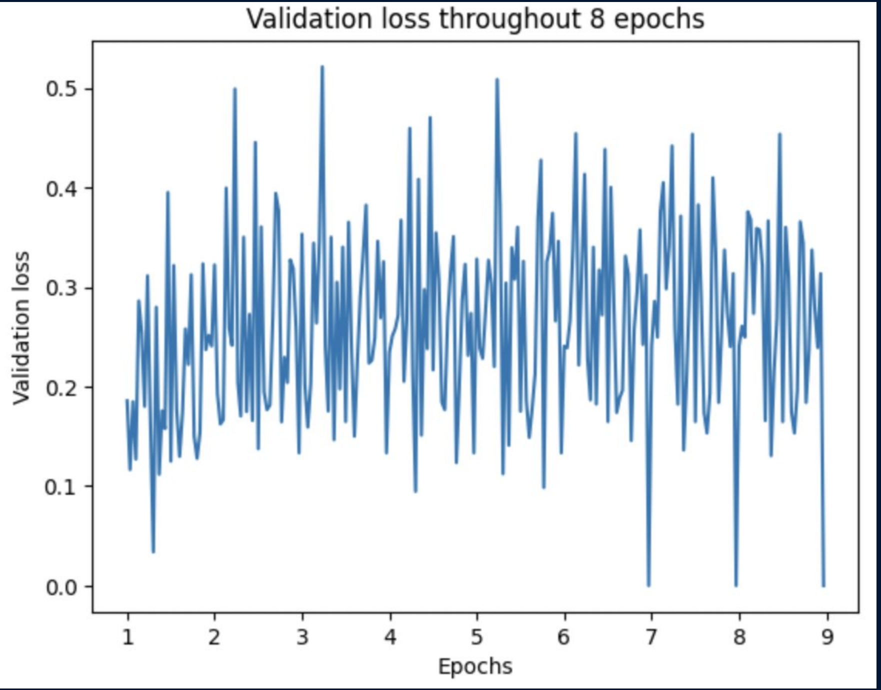
- Currently ~10,000

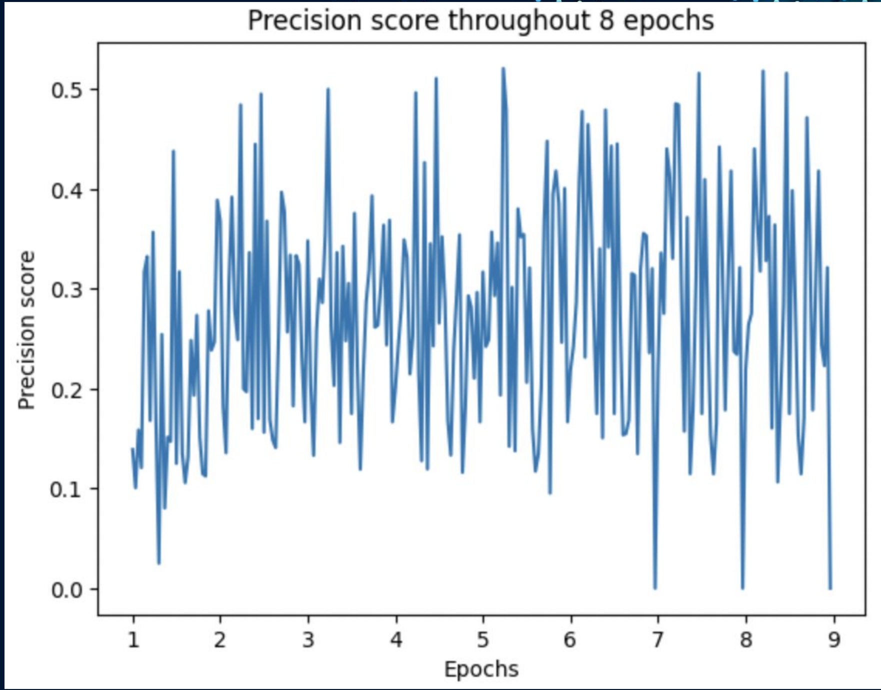
Hyperparameters:
Batch size: 32
Learning rate: 2e-5
Epochs: 8

Optimizer:
AdamW with epsilon of 1e-8

Trained using Google Colab

EVALUATION METRICS





TASK FORMULATION

Generate lyrics based on genre, without sound data

MODEL SELECTION

LSTM

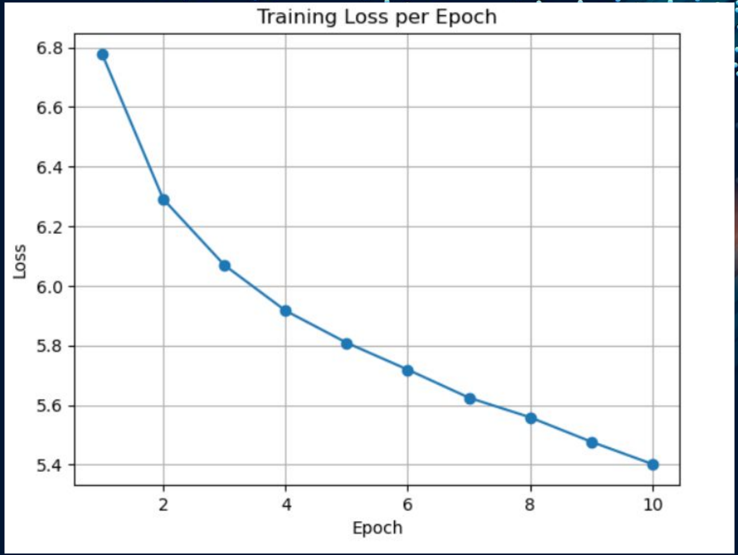
- Prior promising results in “Deep learning in musical lyric generation: An lstm-based approach.” (Gill et. al. 2020)
- Sequences, long term dependencies, light weight!

TRAINING DETAILS

Smaller data set for initial generations
- 4 genres : rock, pop, hiphop, R&B
- Songs: 3468
- Breakdown: 22% rock, 46% R&B, 8% Hiphop, 24% pop

Hyperparameters:
Batch size: 16
Epochs: 10
1 LSTM layer
Teacher Forcing

Optimizer:
Adam with epsilon of 1e-3 for fast convergence on small dataset



GENERATION

Zero-init, data #1
hip hop jazz rock =
<UNK><UNK><UNK><UNK><UNK><UNK><UNK><UNK><UNK><UNK>

Genre-init, data #2
hip hop jazz rock = you 're me , , i , you , i , you , i , you , i , you , i , you , i , you , i , you , i , you , i , you , i , you , i

Genre-init data #3 greedy
hiphop = yeah yeah yeah yeah yeah yeah yeah yeah yeah yeah yeah yeah

GENERATION

Genre-init Data #3 Top-P

HipHop = isnt that leke people sorry that youre talking to paid these 1 girls is mile girls and tryna the day just told me run on now it is the lick the so i i i bottegas and you now all me when i hold on and im fun at my best tryna leave no door do did youre four i and she they can hear it up when im gon the done youre nobody on weed bitch boaw can em make you snow me when i went to done drunk in if you was think no he but nobody else

EVALUATION METRICS

Encoder-Decoder

🎯**Lyrics evaluated by people and metrics**

- Gill et. al. use an LSTM to generate genre-specific lyrics
- Lyrics are not uniform – genres vary *stylistically and structurally*
- They propose **5 linguistic metrics** to evaluate genre adherence

Deep Learning in Musical Lyric Generation: An LSTM-Based Approach

Harrison Gill[1], Daniel (Taesoo) Lee[1], Nick Marwell[1]
[1]*Department of Linguistics, Yale University*

Figure 2: Computed Linguistic Characteristics of Training Data by Genre

Genre	*Average Line Length*	*Song Word Variation*	*Genre Word Variation*	*I vs. You*	*Word Repetition*
Metal	8.98	0.54	0.55	1.35	1.91
Rock	8.48	0.48	0.42	1.48	2.03
Rap	8.17	0.42	0.36	4.71	10.96
Pop	7.33	0.41	0.39	2.25	5.38
Country	8.18	0.48	0.62	1.43	1.52
Jazz	7.17	0.48	0.36	1.23	1.69

Figure 2: Computed Linguistic Characteristics of Training Data by Genre

EVALUATION METRICS

Encoder-Decoder

(*Our scraped data*)

```
country_new: 187 songs
disco_party_blues: 625 songs
disco_party_blues2: 78 songs
disco_party_blues_new: 628 songs
edm_new: 57 songs
hiphop_jazz_rock: 65 songs
hiphop_new: 105 songs
jazz_new: 125 songs
kpop: 187 songs
latin_new: 13 songs
metal_latin_classic: 595 songs
my_playlist: 18 songs
pop_new: 114 songs
pop_rock_hiphop_rap: 190 songs
pop_rock_hiphop_rap_new: 190 songs
rap_edm_country: 84 songs
rap_new: 104 songs
rb_indie_classical: 542 songs
rb_indie_classical_new: 840 songs
rb_new: 229 songs
rock_new: 578 songs
romantic_playlist: 105 songs
sad_happy: 74 songs
test2_set: 328 songs
test_set: 44 songs
trance_ambient: 8 songs
```

Organized data

Top 5 Songs in Genre: HIPHOP_JAZZ_ROCK

	genre	avg_line_length	word_variation	i_vs_you	repetitions
1575	hiphop_jazz_rock	9.848837	0.203070	-1	78
1576	hiphop_jazz_rock	7.113208	0.251989	15	0
1577	hiphop_jazz_rock	7.081081	0.396947	17	9
1578	hiphop_jazz_rock	5.606061	0.443243	15	3
1579	hiphop_jazz_rock	8.650000	0.497110	3	1

Top 5 Songs in Genre: HIPHOP_NEW

	genre	avg_line_length	word_variation	i_vs_you	repetitions
1640	hiphop_new	11.803922	0.433555	9	16
1641	hiphop_new	10.952381	0.363768	-9	7
1642	hiphop_new	8.329412	0.539548	0	10
1643	hiphop_new	7.693878	0.267905	10	11
1644	hiphop_new	7.222222	0.364103	2	24

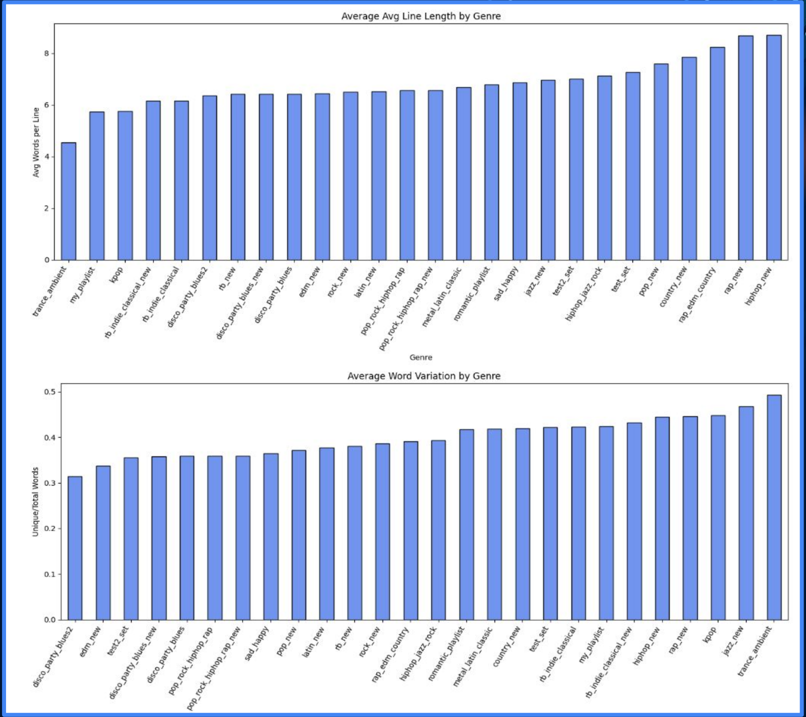
Top 5 Songs in Genre: HIPHOP_NEW

	genre	avg_line_length	word_variation	i_vs_you	repetitions
1640	hiphop_new	11.803922	0.433555	9	16
1641	hiphop_new	10.952381	0.363768	-9	7
1642	hiphop_new	8.329412	0.539548	0	10
1643	hiphop_new	7.693878	0.267905	10	11
1644	hiphop_new	7.222222	0.364103	2	24

Top 5 Songs in Genre: JAZZ_NEW

	genre	avg_line_length	word_variation	i_vs_you	repetitions
1745	jazz_new	3.862069	0.437500	7	0
1746	jazz_new	5.823529	0.646465	4	0
1747	jazz_new	7.222222	0.584615	1	0
1748	jazz_new	5.617647	0.314136	-4	0
1749	jazz_new	5.071429	0.577465	-1	0

Pandas Dataframes



Pyplot!

EVALUATION METRICS

Encoder-Decoder

🎯**Comparison of real vs. generated lyrics**

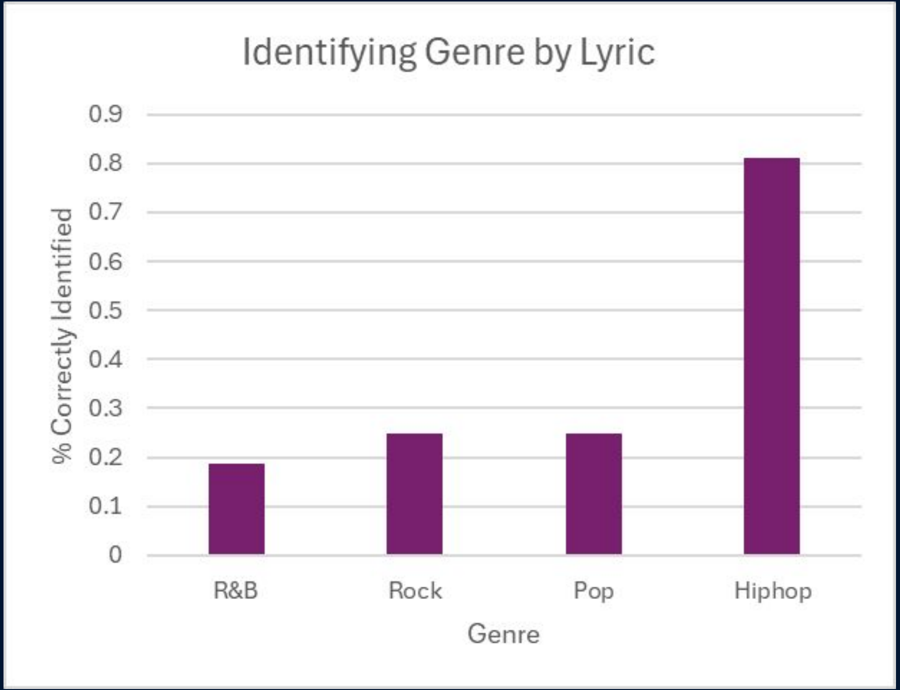
- Gill et al.'s cosine similarity test
- Quantitatively measure **style similarity**
-

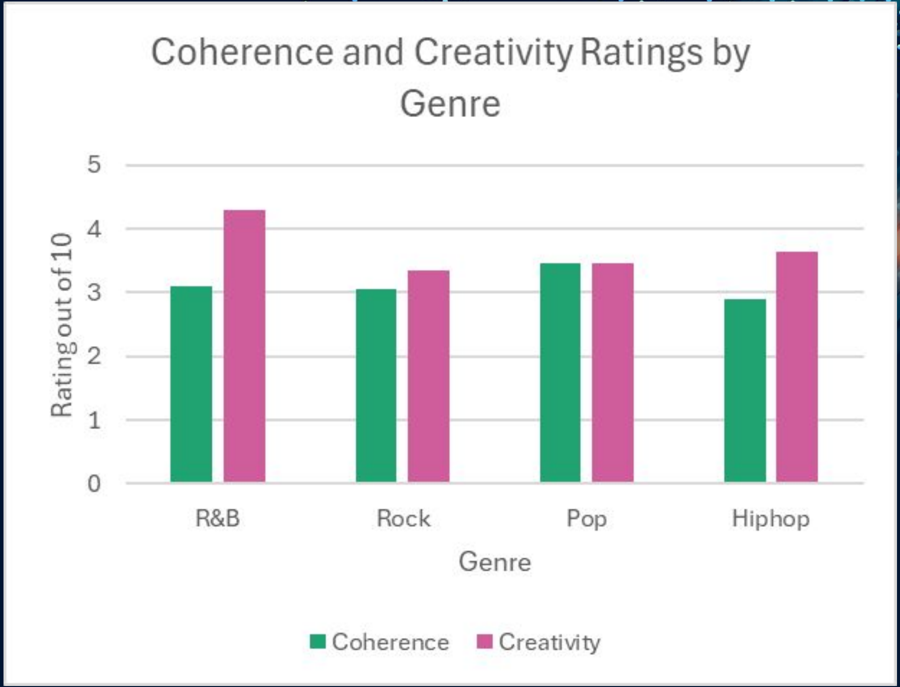
Figure 5: Percentage Changes from original to generated lyrics for each metric and genre

Genre	*Average Line Length*	*Song Word Variation*	*Genre Word Variation*	*I vs. You*	*Word Repetition*
Metal	-54.3	-25.1	-13.5	82.7	88.2
Rock	-38.4	-5.0	-5.9	77.0	34.9
Rap	-52.3	-64.6	-99.0	75.5	-49.6
Pop	-28.0	-16.7	-44.6	94.6	36.5
Country	-74.1	-36.6	-23.0	-229.6	75.2
Jazz	-24.1	-38.7	-78.5	73.7	94.6

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

Human Evaluation of LSTM Lyrics



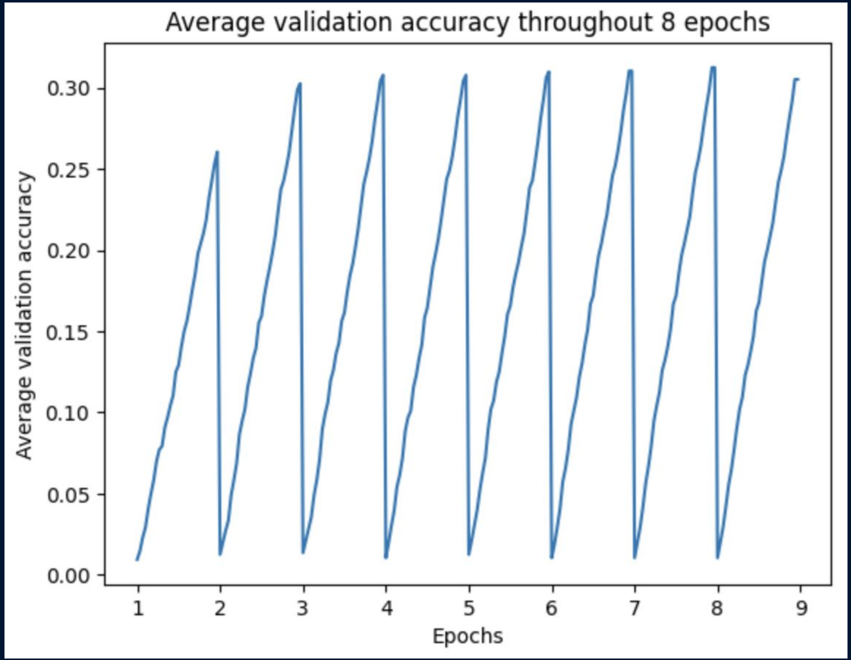


Next steps LSTM

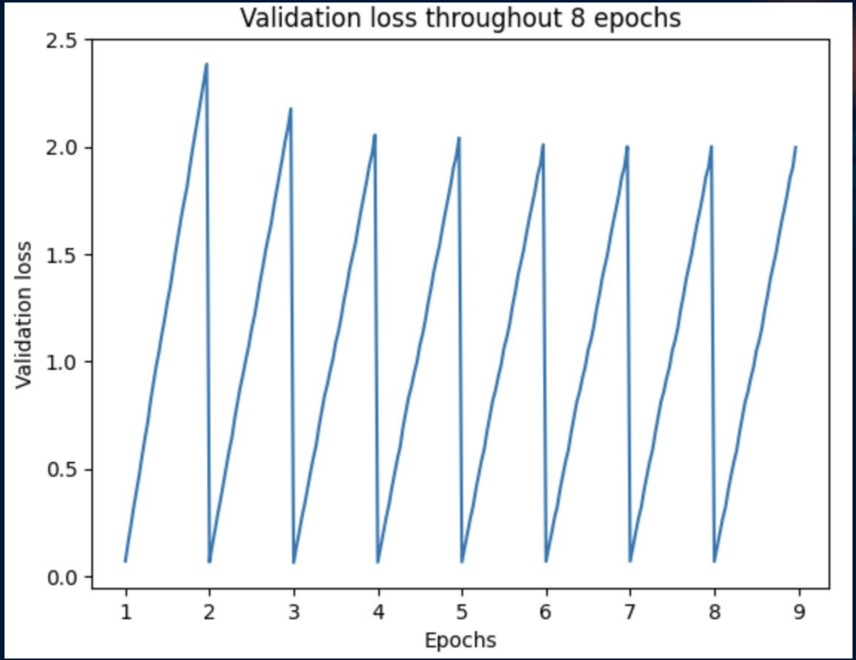
- MORE DATA
- More than 1 LSTM layer, introduce dropout
- Introduce temperature to top-p sampling
- More genres (limited)
- learned embedding layer from the
  fine-tuned BERT model as init hidden state?

BERT RESULTS…so far

Average validation accuracy



Validation loss



ERROR ANALYSIS

BERT may have potential issues with multiple languages. There are some songs that are in different languages that may be causing some errors in the training process. We are working on separating the songs into different languages and try training on solely one language songs.

Another error may be the small batch size. We haven't been able to increase the batch size due to not enough memory on Google Colab.

THANK YOU