

SONG LYRICS GENERATOR USING GPT-2

By: Joshua Mizraji

AGENDA

- 1. Introduction**
- 2. Data Collection**
- 3. Exploratory Data Analysis**
- 4. Modeling**
- 5. Demo**
- 6. Conclusions**

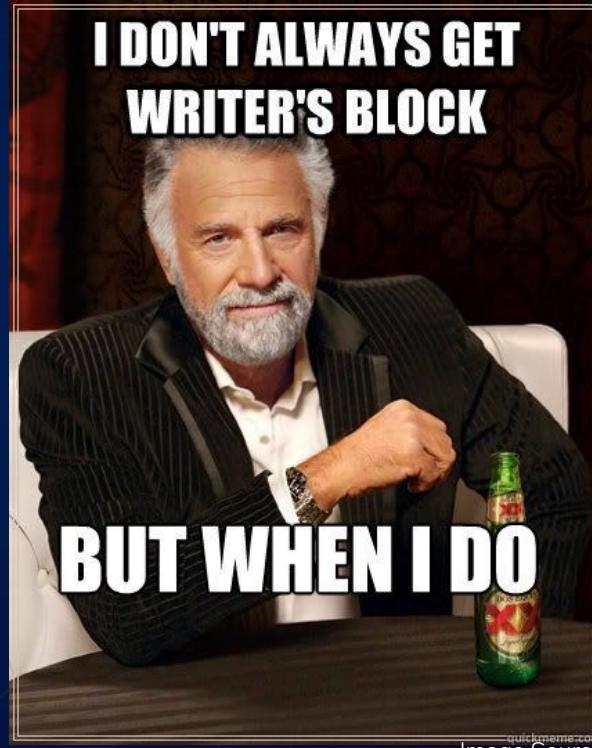
The background features a dark blue gradient with a subtle radial effect. Overlaid on this are several sets of thin, curved lines in white, yellow, and light green. These lines form organic, flowing shapes that resemble waves or stylized leaves. Small, isolated dark grey dots are scattered across the surface.

01 INTRODUCTION

PROBLEM

Writing song lyrics can be a difficult task for musicians because writer's block is hard to overcome and "creative inspiration" is not always something that can be forced.

This leads to musicians feeling frustrated and "stuck" when trying to write song lyrics.



OBJECTIVE

**CREATE A TEXT
GENERATOR THAT MAKES
WRITING SONG LYRICS
EASIER FOR ARTISTS**

SUCCESS CRITERIA

SUGGEST LYRICS

Can I create a text generator that suggests the next few words of a lyric?



ARTIST SPECIFIC

Can I train a model to generate text that sounds like a specific artist?



DYNAMIC OUTPUTS

Can I create an interactive application that allows the user to manipulate the model's output?

The background features a dark blue gradient with abstract white wavy lines and small white dots. On the left side, there is a prominent, multi-layered wavy pattern in yellow, green, and blue, which serves as a visual metaphor for data collection.

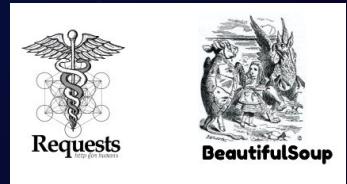
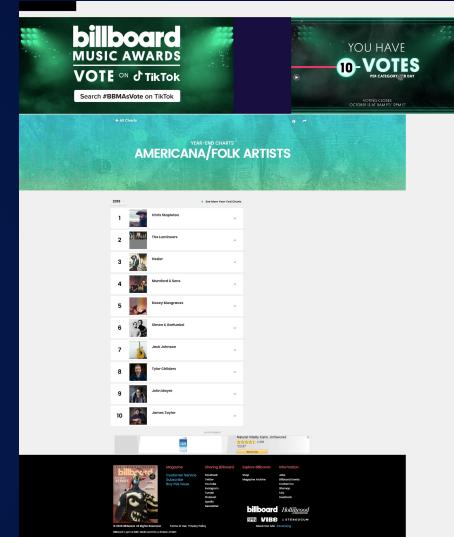
02 DATA COLLECTION

DATA COLLECTION

Billboard Top Artists

- Folk/Americana & Pop
- 2017 - 2019

Data was scraped using
BeautifulSoup and Requests
libraries.



DATA COLLECTION

Lyrics

- Scraped over 5000 song lyrics from the top billboard artists
- Removed “Live Songs” and “Covers” from the dataset

Libraries

- [LyricsGenius](#)
- [Genius API](#)



[Image Source](#)



[Image Source](#)

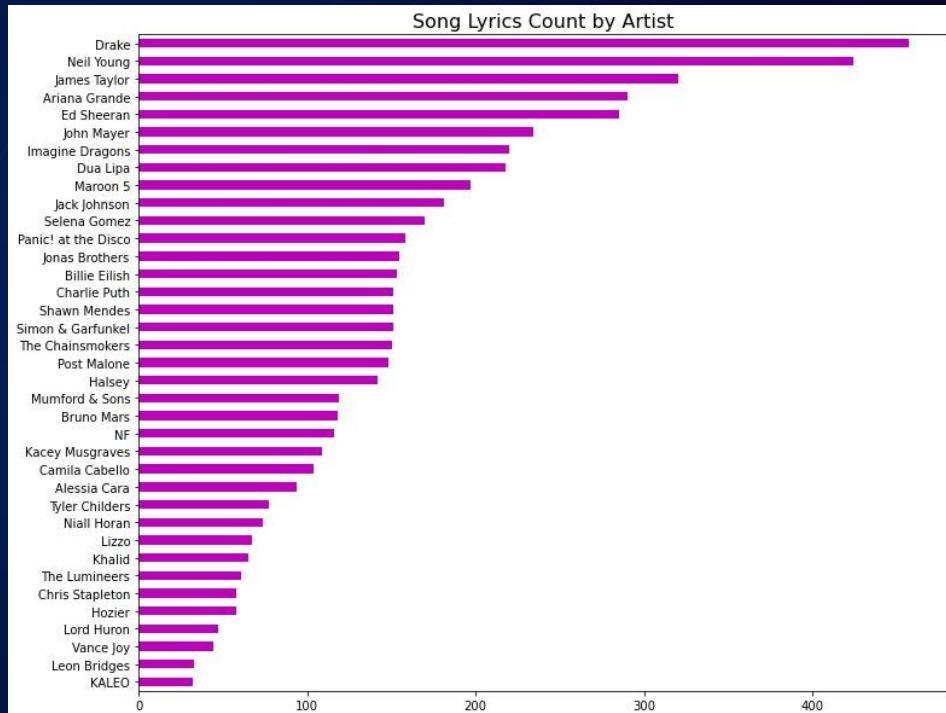
END RESULTS

Lyrics

- Scraped over 5000 song lyrics from the top billboard artists
- Removed “Live Songs” and “Covers” from the dataset

Libraries

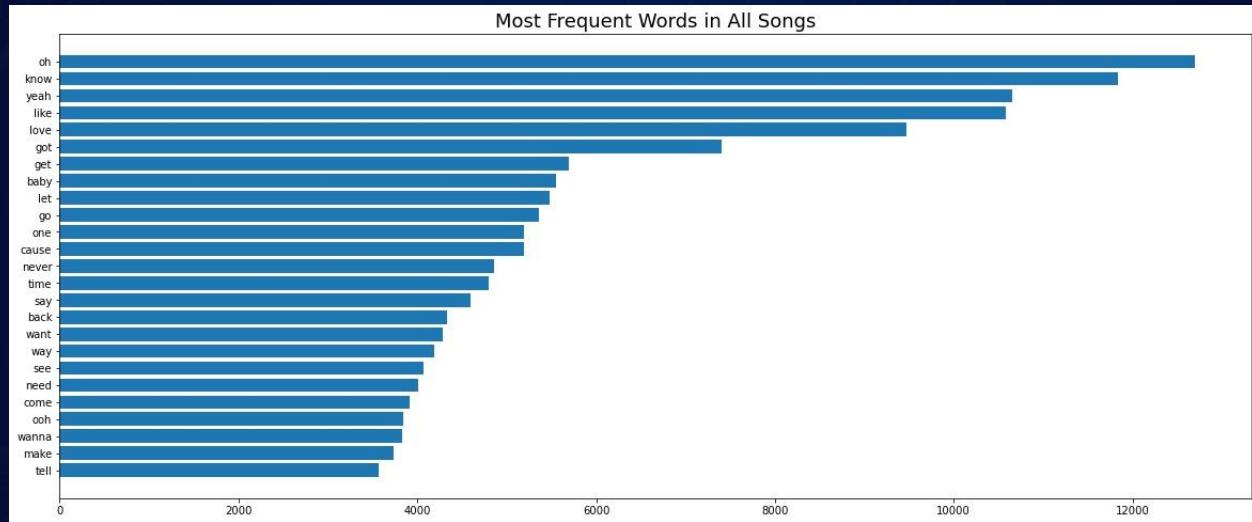
- [LyricsGenius](#)
- [Genius API](#)



The background features a dark blue gradient with a subtle radial blur effect. Overlaid on this are several thin, light-colored wavy lines in shades of yellow, green, and blue. There are also a few small, semi-transparent dark grey circular dots scattered across the surface.

03 EXPLORATORY DATA ANALYSIS

MOST FREQUENT WORDS IN TOP POP & FOLK/AMERICANA SONGS



UNIQUE WORDS
TO
FOLK/AMERICANA

'feel', 'man', 'take', 'well'

UNIQUE WORDS
TO POP

'make', 'ooh', 'tell', 'wanna'

COMPARING ARTIST SIMILARITY BASED ON LYRICS

MOST SIMILAR ARTISTS

COMPARISON	SIMILARITY SCORE
Maroon 5 compared to Shawn Mendes	0.999329
Maroon 5 compared to Jonas Brothers	0.999054
Shawn Mendes compared to The Chainsmokers	0.999037

LEAST SIMILAR ARTISTS

COMPARISON	SIMILARITY SCORE
Post Malone compared to Neil Young	0.975826
Neil Young compared to Lizzo	0.979234
Post Malone compared to Mumford & Sons	0.979630

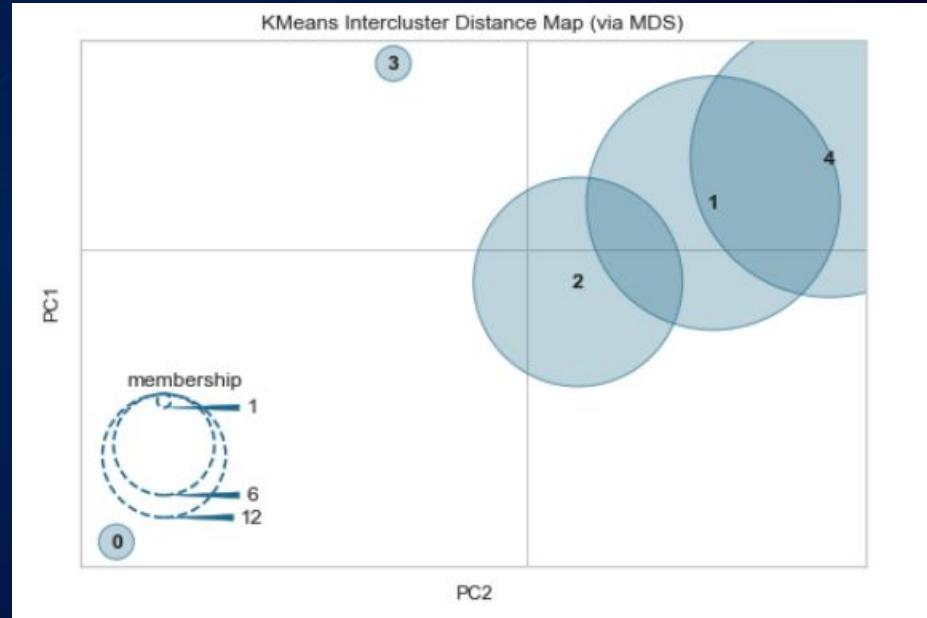
K-MEANS CLUSTERING

K-Means

- Removed Stopwords
- CountVectorizer()
- N-Gram Range (1,3)

Takeaways

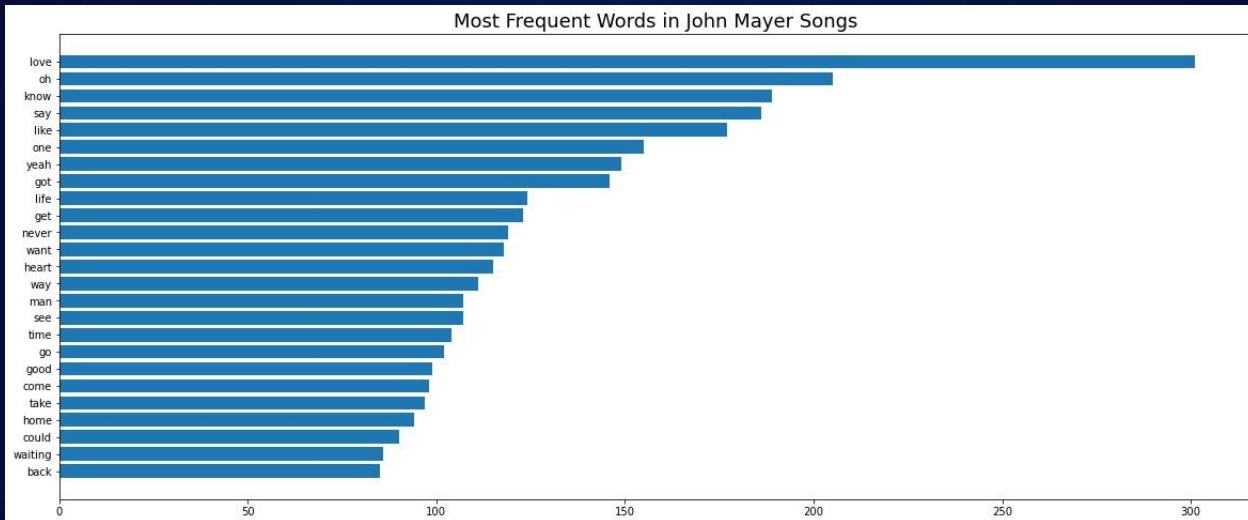
- Drake (0) and Selena Gomez (3) have lyrics that are by the model standards, different from the rest of the artists



The background features a dark blue gradient with a complex pattern of thin, light-colored wavy lines. These lines form several large, sweeping curves on the left side of the frame, transitioning from yellow to green and blue. Small, isolated dark grey dots are scattered across the dark background.

ADDITIONAL EDA: JOHN MAYER

MOST FREQUENT WORDS IN JOHN MAYER SONGS



NOTEABLE WORDS

'love', 'say', 'heart', 'life', 'man', 'good',
'home'

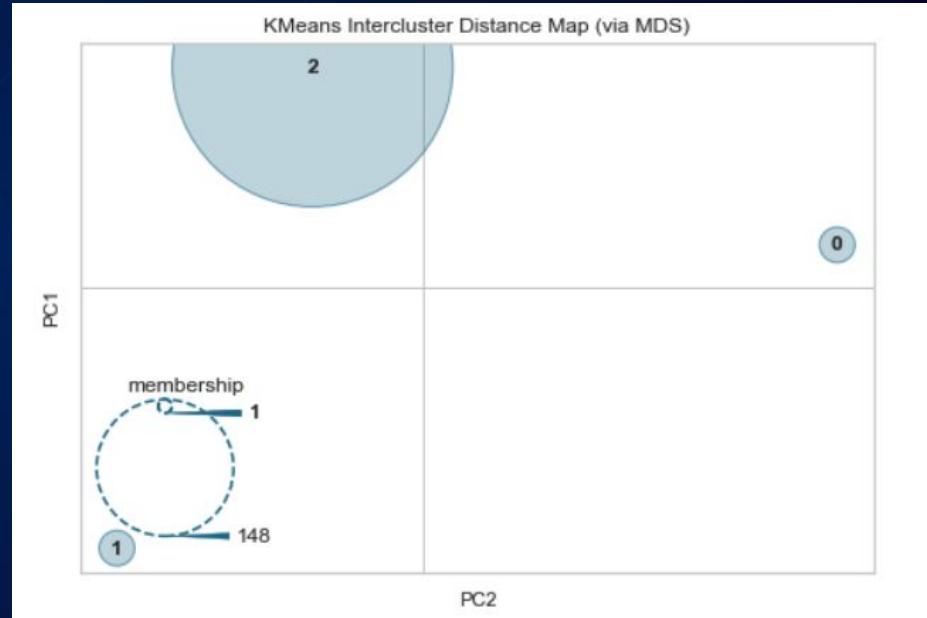
K-MEANS CLUSTERING - JOHN MAYER ANALYSIS

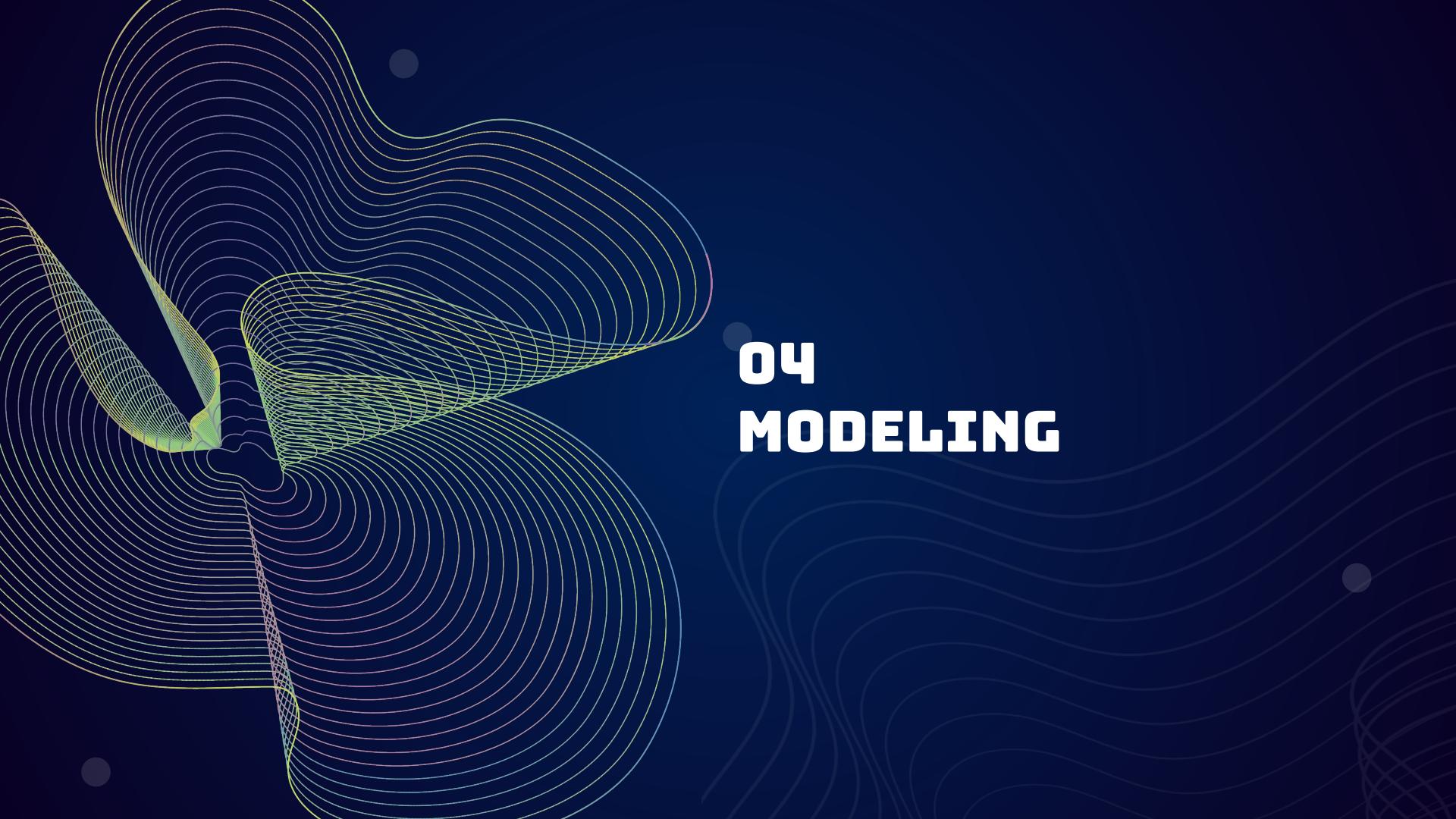
K-Means

- Removed Stopwords
- CountVectorizer()
- N-Gram Range (1,3)

Takeaways

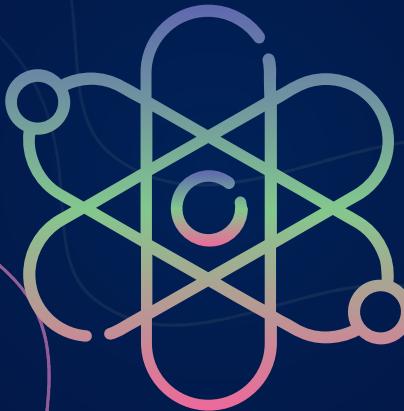
- "Wanna Be Startin' Somethin'"(1) is not an original song
- "Say" (0) is very repetitive with most of the song lyrics “say what you need to say”



The background features a dark blue gradient with abstract white wavy lines and small white dots. On the left side, there is a prominent, multi-layered wavy pattern composed of thin white lines, with a central area highlighted in yellow and green.

04 MODELING

GPT-2



WHAT IS IT?

Text Generation Model built by
[OpenAI](#)

WHY DOES IT EXIST?

To generate the next word,
sentence, or paragraph

WHY IS IT SPECIAL?

Trained on 8 million web pages
with 1.5 billion parameters

HOW DO I USE IT?

Train the model on raw text and
use the parameters to tune the
model to generate the desired
output

MODELING WORKFLOW



GENERATE .TXT FILE

Generate a .txt file with an artist's song lyrics



TRAIN MODEL

Using Google Colab, utilize the GPU power of Google Cloud to train the model on 1000 epochs

DOWNLOAD MODEL

Each model is ~500MB

TEST & TUNE

Using the custom Streamlit app, test and tune the output of the model

The background features a dark blue gradient with a subtle radial effect. Overlaid on this are several thin, curved lines in various colors (yellow, green, blue, pink) that form a complex, flowing pattern. There are also a few small, semi-transparent grey dots scattered across the surface.

05 DEMO

STREAMLIT APP



TUNING PARAMETERS

MODEL SELECTION

LYRICS TEXT BOX

GENERATED TEXT

HUMAN VS AI SCORE

Song Lyrics Generator Using GPT-2

This is a demo of a text generation model trained with GPT-2 to generate song lyrics that sound like John Mayer, Jack Johnson, and Shawn Mendes.

Select an artist to write like:

- John Mayer
- Jack Johnson
- Shawn Mendes

Start writing lyrics...

I've been feeling the same way
It's hard to explain
I once was a hit
Now I'm

Remember to press command + ENTER to save text before generating lyrics

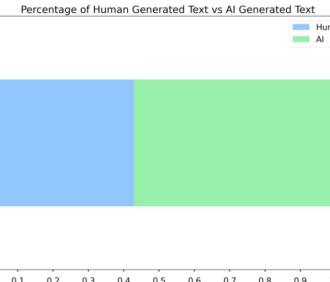
Generate Lyrics

Generating lyrics...

```
+ [0 : "just like many"  
1 : "gones"  
2 : "lying on a"  
3 : "much better"  
4 : "running and waiting"  
5 : "a puddle"  
6 : "no longer"  
7 : "back"  
8 : "getting past the"  
]
```

Human Generated Text: 0.429

AI Generated Text: 0.571



The background features a dark blue gradient with abstract white wavy line patterns. A cluster of wavy lines in the upper left corner is highlighted with a translucent green overlay. Small white dots are scattered across the background.

06 CONCLUSIONS

MAIN TAKEAWAYS

- Using GPT-2, it is entirely possible to create a lyrics generator trained by a specific artist with dynamic outputs
- Even a few suggestions can help remove writer's block and spark "creativity"
- GPT-2 is extremely powerful, however, note the tradeoff that that it is very black box and computationally intensive
- Creating interactive apps is an excellent way to showcase the different parameters of a model and help interested people become familiar

FUTURE WORK

- Continue to build a suite of tools to help artists write more creative song lyrics
- Tune models based on more artists and genres
- Create and compare a model using Google's BERT
- Deploy to production using Heroku
- Gather feedback and feature requests from other artists and developers in the community

THANK YOU!

Do you have any questions?

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.

Please keep this slide for attribution.