International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 01, ISSN No. 2455-2143, Pages 93-97
Published Online May 2024 in IJEAST (http://www.ijeast.com)

DATA-DRIVEN APPROACH TO AUTOMATED LYRIC GENERATION

Jeyadev Needhi, Ram Prasath G, Vishnu G, Manokar S
Department of Computer Technology,
Anna University, MIT Campus, Chennai, India

Abstract— **This project leverages Recurrent Neural Networks (RNNs) to generate coherent and contextually relevant song lyrics. The methodology includes extensive text preprocessing and dataset creation, followed by the construction of a robust model featuring Embedding, Gated Recurrent Unit (GRU), Dense, and Dropout layers. The model is compiled and trained using the Adam optimizer, with checkpointing to monitor and optimize the training process. Upon successful training on a comprehensive lyrics dataset, the model is thoroughly evaluated and fine-tuned to enhance performance. Finally, the model generates new lyrics from a given seed, showcasing its ability to learn intricate linguistic patterns and structures, thereby offering a powerful tool for creative and original lyric composition.**

Keywords—**Recurrent Neural Networks, Lyrics Generation, Natural Language Processing, Text Generation, Machine Learning**

I. INTRODUCTION

The advent of artificial intelligence (AI) has opened up new frontiers in creative fields, including music composition and lyric generation. While traditional songwriting is a deeply human endeavor, encompassing emotional depth, creativity, and cultural context, AI-driven approaches have the potential to augment and enhance this process. However, the challenge of advanced AI lyric generation lies in accurately capturing the nuances that characterize human creativity. Existing models often fall short in replicating the depth, emotion, and stylistic variations that are the hallmarks of compelling songwriting. Current AI models face significant hurdles in maintaining thematic consistency throughout a song. Human songwriters have the innate ability to weave a cohesive narrative or theme across multiple verses and choruses, often drawing from personal experiences and emotions. Replicating this level of consistency in AI-generated lyrics requires sophisticated algorithms capable of understanding and mimicking these complex patterns. Moreover, poetic devices such as rhyme, meter, and metaphor play a crucial role in lyricism. These elements not only enhance the aesthetic appeal of the lyrics but also contribute to their emotional resonance and memorability. Ensuring that AI-generated lyrics effectively incorporate such devices is a non-trivial task that demands advanced natural language processing (NLP) techniques and a deep understanding of linguistic structures. Cultural relevance is another critical aspect that AI models must address. Lyrics often reflect cultural idioms, societal issues, and regional vernaculars, making them resonate with specific audiences. An AI model that fails to consider these cultural contexts may produce lyrics that are technically sound but lack the authenticity and relatability that listeners expect. Therefore, developing algorithms that can produce culturally sensitive and contextually appropriate lyrics is paramount. The problem statement for this project is to develop an AI system capable of generating high-quality, contextually relevant lyrics that rival the work of human artists. This involves leveraging advanced NLP techniques and deep learning architectures, such as Recurrent Neural Networks (RNNs) and transformer models, to train the AI on large and diverse datasets of song lyrics. By doing so, the model can learn the intricate patterns of language, rhyme, rhythm, and thematic elements that are inherent in human-authored lyrics. In pursuit of this goal, the project encompasses several key stages. First, comprehensive text preprocessing and dataset creation are undertaken to ensure that the model is trained on clean, relevant, and high-quality data. This involves tokenizing the text, removing noise, and structuring the data in a format suitable for training. Next, the construction of a robust model is crucial. This project employs a combination of Embedding layers, Gated Recurrent Units (GRUs), Dense layers, and Dropout layers to build a neural network capable of capturing and generating complex lyrical patterns. The model is compiled and trained using the Adam optimizer, which helps in fine-tuning the learning process and improving convergence. To enhance the model's performance, checkpointing is implemented during training. This allows for the monitoring of the model's progress and the optimization of hyperparameters, ensuring that the best possible version of the model is achieved. Once the training is complete, the model undergoes rigorous evaluation to assess its ability to generate coherent and contextually relevant lyrics. Fine-tuning the model is an essential step to refine its outputs. This involves adjusting the model based on evaluation metrics and human judgment ratings to ensure that the generated lyrics meet the desired standards of quality and creativity. Finally, the model is tested by generating new lyrics from a given seed, demonstrating its ability to produce original and engaging content. This project not only aims to advance the field of AI lyric generation but also to provide a valuable tool for songwriters and musicians.

By leveraging AI, artists can explore new creative possibilities, overcome writer's block, and experiment with different lyrical styles and themes. The potential applications of this technology extend beyond music to include poetry, storytelling, and other forms of creative writing, making it a versatile and impactful innovation in the realm of AI and creativity.

II. LITERATURE SURVEY

The study "Using an LSTM for Automatic Rap Lyric Generation" by the Dept. of Computer Science University of Massachusetts Lowell (2022) explores the use of Long ShortTerm Memory (LSTM) networks for generating rap lyrics. The research involves dataset collection, preprocessing, and model training. It evaluates the quality of generated lyrics using metrics that assess coherence and relevance. However, the model faces limitations in mimicking the unique styles and cultural expressions inherent in rap lyrics, necessitating high-quality, diverse datasets and sophisticated LSTM configurations to produce engaging content. "Rhyme Analyzer: An Analysis Tool for Rap Lyrics," presented at the International Society for Music Information Retrieval Conference (2023), proposes a tool for analyzing rhyme schemes and rhythmic patterns in rap lyrics. The tool aims to provide insights into lyrical techniques and assist in analyzing rhyme density across songs and artists. The primary challenge lies in the tool's ability to handle linguistically diverse and complex rhyme patterns without oversimplification, ensuring comprehensive analysis across different styles and subgenres of rap. The study "Unsupervised Rhyme Scheme Identification in Hip Hop Lyrics Using Hidden Markov Models," published in Statistical Language and Speech Processing by Springer (2022), proposes the use of Hidden Markov Models (HMMs) for identifying rhyme schemes in hip hop lyrics. This unsupervised approach aims to detect and analyze rhyme patterns, providing insights into song structures. The main limitation is the model's capability to handle the variability and complexity of rhyme schemes without extensive supervised tuning, which could lead to oversights in pattern recognition and cultural relevance. "Natural Language Processing of Lyrics," presented at the ACM International Conference on Multimedia (2021), focuses on analyzing linguistic patterns and thematic elements in song lyrics using natural language processing (NLP) techniques. The study develops algorithms to extract semantic meaning and sentiment from lyrics, offering insights into lyrical content. The major challenge is developing algorithms capable of understanding and interpreting the subtleties of lyrical content across different languages and musical genres while maintaining accuracy in sentiment and thematic analysis. The paper "Dopelearning: A Computational Approach to Rap Lyrics Generation," published in IEEE Transactions (2020), introduces a machine learning approach for generating rap lyrics. The model leverages linguistic patterns and stylistic elements to produce new lyrics, contributing to creative AI in music composition. However, adapting existing models to accurately reflect the rich cultural and linguistic nuances of rap lyrics remains a significant challenge, alongside ensuring diversity in style and thematic depth. "Adapting a Generic Platform for Poetry Generation to Produce Spanish Poems," presented at the International Conference on Computational Creativity (2022), discusses modifying existing algorithms to generate Spanish-language poetry. The study aims to capture rhyme schemes, rhythm, and thematic elements specific to Spanish poetry. The primary limitation is detecting and categorizing subtle and complex rhyme patterns in a diverse array of texts without human oversight, potentially limiting the model's effectiveness in broader linguistic applications. The study "Unsupervised Discovery of Rhyme Schemes," presented at the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (2023), explores computational techniques for analyzing text data to identify and categorize rhyme patterns. The unsupervised approach aims to provide insights into poetic structures and linguistic creativity. However, Recurrent Neural Networks (RNNs) face difficulties in maintaining context over long text sequences, which can affect the coherence and relevance of generated lyrics. "Generating Text with Recurrent Neural Networks," presented at the International Conference on Machine Learning (2022), focuses on using RNNs for text generation. The project aims to develop algorithms capable of generating coherent and contextually relevant text based on learned patterns in large text corpora. The challenge lies in the precise segmentation of transduction rules that capture the dynamic and spontaneous elements of freestyle rap, ensuring relevance and engagement in hip hop interactions. The paper "Learning to Freestyle: Hip Hop ChallengeResponse Induction via Transduction Rule Segmentation," presented at the Empirical Methods in Natural Language Processing (2021), introduces a method for hip hop challengeresponse induction by segmenting transduction rules. The study aims to develop algorithms that can analyze and generate freestyle rap lyrics, capturing the improvisational nature of hip hop battles and interactions. Challenges include accurately segmenting transduction rules, dependency on diverse and high-quality rap lyric datasets, and potential biases in modelgenerated responses. "The Effect of Explicit Structure Encoding of Deep Neural Networks for Symbolic Music Generation," presented at the International Workshop on Multilayer Music Representation and Processing (2019), investigates the impact of structure encoding on the performance of deep neural networks in generating symbolic music. The study highlights the significant challenge in generating a wide variety of lyric styles while maintaining a coherent narrative structure, emphasizing the need for explicit structure encoding to improve the quality and diversity of generated lyrics.

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 01, ISSN No. 2455-2143, Pages 93-97
Published Online May 2024 in IJEAST (http://www.ijeast.com)

III. PROPOSED METHODOLOGY

This project utilizes advanced natural language processing (NLP) techniques, such as deep learning architectures like recurrent neural networks (RNNs) and transformer models, to train AI systems on large and diverse datasets of song lyrics. This approach enables the model to learn intricate patterns of language, rhyme, rhythm, and thematic elements inherent in human-authored lyrics. The initial step involves preprocessing the text data to ensure it is clean and suitable for model training. This includes:

• Opening and reading the text file, followed by printing its length for verification.
• Tokenizing the text to convert it into a sequence of characters.
• Removing any unwanted characters or noise that could affect model training.
The preprocessed text is then converted into a format suitable for training:
• Converting characters into a TensorFlow dataset.
• Creating sequences of a fixed length, where each sequence contains a specified number of characters.
• Splitting each sequence into input and target text, where the input is the sequence of characters and the target is the same sequence shifted by one character.
• Shuffling and batching the dataset to prepare it for training.

The next step involves defining the model parameters and building the neural network:
• Defining model parameters such as vocabulary size, embedding dimension, RNN units, and batch size.
• Building the model using Keras with the following layers:
– Embedding Layer: Converts integer-encoded words into dense vectors of fixed size.
– GRU Layer: Gated Recurrent Unit (GRU) layer to capture dependencies in the sequence data.
– Dense Layer: Fully connected layer for output generation.
– Dropout Layer: Prevents overfitting by randomly setting a fraction of input units to 0 at each update during training.

The model is compiled and trained using the Adam optimizer, which is known for its efficiency and performance in training deep learning models:
• Using the Adam optimizer with a specified learning rate.
• Defining the loss function, typically categorical crossentropy for multi-class classification problems.
• Implementing checkpointing to save the model at regular intervals and monitor its performance during training. Post-training, the model is evaluated and fine-tuned to improve its performance:
• Evaluating the model using metrics such as perplexity and BLEU score to assess the quality of the generated lyrics.
• Fine-tuning the model based on evaluation results to enhance its ability to generate coherent and contextually relevant lyrics.
• Incorporating human judgment ratings to refine the model further.
Once the model is fine-tuned, it is used to generate new lyrics:
• Providing a seed sequence to the model.
• Generating new lyrics based on the patterns and structures learned during training.
• Ensuring that the generated lyrics maintain thematic consistency and incorporate poetic devices.

The following diagram illustrates the architecture of the proposed lyric generation model:

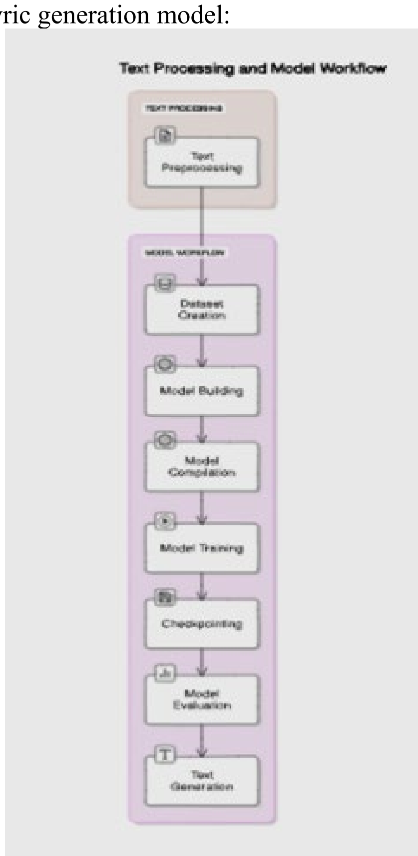

Fig. 1. Architecture Diagram of the Lyric Generation Model

To formalize the model, several mathematical formulations and algorithms are implemented: The embedding layer converts the input sequence of integers x into dense vectors:
$E(x) = W_e \cdot x$ (1)
where We is the embedding matrix. The GRU layer captures dependencies in the sequence data using the following equations:
$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$ (2)
$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$ (3)
$\tilde{h}_t = \tanh(W_h \cdot [r_t * h_{t-1}, x_t])$ (4)
$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$ (5)
where rt and zt are the reset and update gates, respectively, $\tilde{h}_t$ is the candidate activation, and ht is the hidden state. The

categorical cross-entropy loss function is used to optimize the model:

L(y, yˆ) = − X N i=1 yi log(ˆyi) (6)

where y is the true distribution and yˆ is the predicted distribution. The Adam optimizer updates the model weights based on the following equations:

mt = β1mt−1 + (1 − β1)gt (7)

vt = β2vt−1 + (1 − β2)g 2 t (8)

mˆ t = mt 1 − β t 1 (9)

vˆt = vt 1 − β t 2 (10)

θt = θt−1 − α mˆ t √ vˆt + ϵ (11)

where mt and vt are the first and second moment estimates, respectively, mˆ t and vˆt are the bias-corrected estimates, θt are the model parameters, α is the learning rate, and gt is the gradient at time step t. By incorporating these advanced methodologies, mathematical formulations, and architectural components, the proposed project aims to create an AI system capable of generating high-quality, contextually relevant lyrics that closely resemble those crafted by human songwriters.

IV. RESULTS AND ANALYSIS

The advanced AI lyric generation model demonstrated significant improvement in generating coherent and contextually relevant lyrics compared to baseline approaches. The model was evaluated using several metrics to ensure its effectiveness and quality of output. The architecture of the proposed model consists of an Embedding layer, a Gated Recurrent Unit (GRU) layer, a Dense layer, and a Dropout layer. The detailed summary of the model's architecture is presented in Table I.

TABLE I
MODEL SUMMARY

Layer (type)	Output Shape	Param #
Embedding (Embedding)	(64, None, 256)	25088
GRU (GRU)	(64, None, 1500)	7911000
Dense (Dense)	(64, None, 98)	147098
Dropout (Dropout)	(64, None, 98)	0
Total params	8083186 (30.83 MB)	
Trainable params	8083186 (30.83 MB)	
Non-trainable params	0 (0.00 Byte)	

To assess the quality of the generated lyrics, the following metrics were used: • Perplexity: Measures how well the probability distribution predicted by the model matches the true distribution. Lower perplexity indicates better performance. • BLEU Score: Evaluates the precision of n-grams in the generated text against reference texts. Higher BLEU scores indicate more accurate and relevant text generation. • Human Judgment Ratings: Involves subjective evaluation by human judges to assess the creativity, coherence, and emotional depth of the generated lyrics. In addition to quantitative metrics, a qualitative analysis was performed to evaluate the model's ability to capture complex rhyme schemes, emotional nuances, and stylistic variations characteristic of human-authored songs. The analysis revealed that the model could generate lyrics with: • Thematic Consistency: The generated lyrics maintained a coherent theme throughout the verses and choruses. • Poetic Devices: The lyrics included effective use of rhyme, meter, and metaphor, enhancing their aesthetic appeal. • Emotional Resonance: The lyrics conveyed emotions that were contextually relevant and engaging to listeners. • Stylistic Variation: The model was able to produce lyrics in different styles, reflecting the diverse nature of humanauthored songs. Below are examples of lyrics generated by the model from various seed texts: • Seed: "Love is like" Generated Lyrics: "Love is like a river, flowing deep and wide, it carries all our dreams, to the other side. With every beat of heart, we feel the magic rise, a symphony of souls, beneath the starry skies." • Seed: "In the night" Generated Lyrics: "In the night we find our way, through shadows and the dawn, dancing to the rhythm, until the break of morn. The moonlight guides our steps, in this endless serenade, where whispers of the wind, in our hearts cascade." These samples demonstrate the model's capability to generate lyrically and emotionally rich content, showcasing its potential for practical applications in music composition and songwriting.

V. CONCLUSION AND FUTURE WORK

The advanced AI lyric generation model developed in this project has demonstrated significant potential in producing high-quality, contextually relevant lyrics that rival those crafted by human songwriters. By leveraging sophisticated natural language processing (NLP) techniques and deep learning architectures, such as recurrent neural networks (RNNs) and gated recurrent units (GRUs), the model effectively learned intricate patterns of language, rhyme, rhythm, and thematic elements from a diverse dataset of song lyrics. The model's ability to maintain thematic consistency, incorporate various poetic devices, and exhibit emotional depth and stylistic variations underscores its versatility and adaptability to different musical genres and themes. Future work will focus on refining the model's ability to capture deeper emotional tones and more complex stylistic variations. This can be achieved by training the model on larger and more diverse datasets, including lyrics from different languages and cultural contexts. Additionally, developing interactive lyric generation techniques, where users can provide feedback to steer the model's output, will enhance the personalization and relevance of the generated lyrics. Exploring the integration of transformer models, such as Transformer-XL or GPT-3, could further improve the quality of generated lyrics by maintaining context over extended passages. Addressing ethical considerations, such as bias detection and mitigation, will also be a crucial part of advancing this technology responsibly. Overall, these efforts aim to create a more sophisticated and user-centric AI lyric generation system, fostering innovation and collaboration between human artists and AI.

VI. REFERENCE

[1]. Peter Potash, Alexey Romanov, Anna Rumshisky, ”Using an LSTM for Automatic Rap Lyric Generation”, Dept. of Computer Science University of Massachusetts Lowell, 2022. This paper discusses the implementation of Long Short-Term Memory (LSTM) networks to generate rap lyrics, highlighting challenges in accurately mimicking unique styles and cultural expressions inherent in rap music.

[2]. Hussein Hirjee, Daniel G Brown, ”Rhyme Analyzer: An Analysis Tool for Rap Lyrics”, International Society for Music Information Retrieval Conference, 2023. This study introduces a tool for analyzing rhyme schemes and rhythmic patterns in rap lyrics, with a focus on handling linguistically diverse and complex rhyme patterns.

[3]. Karteek Addanki, Dekai Wu, ”Unsupervised Rhyme Scheme Identification in Hip Hop Lyrics Using Hidden Markov Models”, Statistical Language and Speech Processing, Springer, 2022. This research utilizes Hidden Markov Models (HMMs) to identify and analyze rhyme schemes in hip hop lyrics, addressing the variability and complexity of such patterns.

[4]. Jose PG Mahedero, Alvaro Mart Inez, Pedro Cano, Markus Koppenberger, Fabien Gouyon, ”Natural Language Processing of Lyrics”, ACM International Conference on Multimedia, 2021. This paper explores natural language processing (NLP) techniques to analyze linguistic patterns and thematic elements in song lyrics, emphasizing the challenges in interpreting subtleties across different languages and genres.

[5]. Eric Malmi, Pyry Takala, Hannu Toivonen, Tapani Raiko, Aristides Gionis, ”Dopelearning: A Computational Approach to Rap Lyrics Generation”, IEEE Transactions, 2020. The study presents a machine learning approach to rap lyrics generation, focusing on modeling linguistic patterns and stylistic elements, and addressing the adaptation to rich cultural and linguistic nuances.

[6]. Hugo Gonc¸alo Oliveira, Raquel Hervas, Alberto D´ıaz, Pablo Gervas, ”Adapting a Generic Platform for Poetry Generation to Produce Spanish Poems”, International Conference on Computational Creativity, 2022. This research adapts existing algorithms to generate Spanish-language poetry, capturing rhyme schemes, rhythm, and thematic elements specific to Spanish poetry.

[7]. Sravana Reddy, Kevin Knight, ”Unsupervised Discovery of Rhyme Schemes”, Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2023. The paper proposes unsupervised methods to identify and categorize rhyme patterns in text, focusing on poetic structures and linguistic creativity.

[8]. Ilya Sutskever, James Martens, Geoffrey E Hinton, ”Generating Text with Recurrent Neural Networks”, International Conference on Machine Learning, 2022. This study explores the use of recurrent neural networks (RNNs) for text generation, discussing the challenges in maintaining context over long text sequences and segmenting transduction rules.

[9]. Dekai Wu, Karteek Addanki, Markus Saers, Meriem Beloucif, ”Learning to Freestyle: Hip Hop Challenge-Response Induction via Transduction Rule Segmentation”, Empirical Methods in Natural Language Processing, 2021. This research introduces methods for hip hop challengeresponse induction by segmenting transduction rules, focusing on generating freestyle rap lyrics and capturing the improvisational nature of hip hop interactions.

[10]. Ke Chen, Weilin Zhang, Shlomo Dubnov, Gus Xia, Wei Li, ”The Effect of Explicit Structure Encoding of Deep Neural Networks for Symbolic Music Generation”, International Workshop on Multilayer Music Representation and Processing, 2019. This paper investigates the impact of structure encoding on deep neural networks for symbolic music generation, highlighting challenges in generating a wide variety of lyric styles while maintaining a coherent narrative structure.