

GenMelody Automatic Melody Generation For A Given Phrase

Ramadevi Jammu
 DeptArtificial Intelligence & Machine
 Learning
 Anil Neerukonda Institute Of
 Technology And Sciences
 Visakhapatnam, India
<https://orcid.org/0009-0009-6579-3236>
 jammuramadevi12345@gmail.com

Thogaru Mallika
 Artificial Intelligence & Machine
 Learning
 Anil Neerukonda Institute Of
 Technology And Sciences
 Visakhapatnam, India
<https://orcid.org/0009-0009-2559-1402> mallika.csd@anits.edu.in

Abstract—Melody song generation for a given phrase is a crucial task in symbolic music composition and computational creativity. This paper presents GenMelody, an automatic melody song generation system for different Indian languages designed to produce coherent and expressive melodies for input text phrases. The study explores various approaches, including rule-based systems, statistical models, and deep learning-based techniques. We analyze the effectiveness of Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Transformer-based architectures in melody synthesis. Additionally, we discuss hybrid approaches that integrate music theory constraints with AI-generated compositions. The experimental results highlight the advantages and challenges of different methods, providing insights into future research directions in AI-driven music generation.

Keywords—Melody song Generation, AI Music Composition, Deep Learning, Transformer Models, Symbolic Music, Hybrid Approaches.

I. INTRODUCTION

Melody song generation has been a long-standing challenge in computational music composition, requiring an understanding of harmonic structures, rhythm, and expressive variations. While traditional rule-based approaches relied on predefined constraints, machine learning techniques have enabled more creative and diverse musical outputs. This paper introduces GenMelody, an AI-driven system designed to generate melody songs in different Indian languages for a given phrase. The system leverages deep learning architectures and explores various melody generation strategies. The objective is to analyze the strengths and weaknesses of different models and propose a hybrid approach that balances structure with creativity.

II. LITERATURE SURVEY

The intersection of music, human development, and artificial intelligence has been widely explored across diverse research domains. Studies like those by Welch et al. [1] and Hallam [2] emphasize music's critical role in enhancing cognitive, emotional, and social well-being across the lifespan. As AI advances, researchers have leveraged machine learning for automated lyric and music generation. Chen and Lerch [3] utilized SeqGANs for melody-conditioned lyric generation, while Kashyap et al. [4] and Dhandapani et al. [6] applied LSTM and RNN

models to generate lyrics aligned with emotional and thematic content. Barbieri et al. [5] introduced Markov constraints for stylistically consistent lyrics, and Vechtomova et al. [7] enhanced lyric generation using VAEs and artist-specific embeddings. Broader music generation systems have incorporated NLP techniques to create melodies from lyrics, as highlighted by Wang et al. [8]. Traditional rule-based systems [9], Markov models [10], and deep learning methods such as WaveNet [11] and Musenet [12] have significantly contributed to the generation of structured, expressive music. Moreover, platforms like Magenta [13] offer open-source tools supporting a variety of generative models, encouraging collaboration between AI and human creativity. These studies collectively inform and inspire the development of AI-driven, multilingual, and melody-conditioned song generation systems

III. RELATED WORK

Melody song generation has been explored through a range of approaches, evolving from symbolic, rule-based systems to modern deep learning architectures. Early rule-based systems strictly followed predefined music theory constraints, such as scales, rhythm patterns, and harmonic rules. While these systems ensured structural correctness, they often lacked creative variability and adaptability to different musical contexts.

Statistical models, including Markov Chains and Hidden Markov Models (HMMs), introduced probabilistic frameworks for modeling note transitions. These methods were able to capture local dependencies within melodies but struggled with maintaining coherence across longer sequences, often producing monotonous or disjointed outputs.

The emergence of deep learning brought significant advancements to melody generation. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks enabled systems to learn from sequential data, resulting in smoother melodic transitions and more expressive outputs. However, these models were still limited by issues such as vanishing gradients and difficulties in modeling long-term dependencies.

More recently, Transformer-based architectures have demonstrated substantial improvements in generating coherent and stylistically rich melodies. Models like Music Transformer and MuseNet leverage self-attention mechanisms to capture long-range dependencies, enabling them to produce more musically engaging sequences. Despite their success, these models often require large-scale training data and computational resources, and they may lack cultural specificity when applied to diverse musical traditions.

Hybrid models that integrate rule-based constraints with neural network outputs have also gained attention. These systems aim to combine the structure and musicality of rule-based methods with the creative flexibility of AI-driven generation, striking a balance between control and innovation.

IV. EXISTING SYSTEM

a) Melody generation has been approached through various computational techniques, each with its own strengths and limitations.

b) A. Rule-Based Systems

c) Rule-based systems rely on predefined music theory principles, such as scales, chord progressions, and rhythmic patterns, to construct melodies. While these systems ensure harmonic correctness and structural validity, they often lack the flexibility to produce creative or emotionally expressive compositions. As a result, their outputs can appear repetitive and rigid.

d) B. Statistical Models (Markov Chains, HMMs)

e) Statistical models such as Markov Chains and Hidden Markov Models (HMMs) use probabilistic transitions between notes to generate melodies. These models capture local dependencies effectively but struggle to model long-term musical structures, often resulting in fragmented or monotonous sequences that lack coherent phrasing.

f) C. Machine Learning-Based Models

g) Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks introduced the ability to learn from musical sequences, enhancing melodic flow and expression. However, issues like vanishing gradients and limited memory capacity hinder their performance on longer musical phrases. These models also require large and diverse training datasets to generalize effectively across musical genres.

h) D. Transformer-Based Models

i) Transformer architectures address the limitations of sequential models by leveraging self-attention mechanisms to model long-range dependencies. Pretrained models such as Music Transformer and MuseNet have shown strong performance in generating coherent and expressive melodies. Despite their advantages, these models are computationally intensive, making them less suitable for real-time applications or low-resource environments.

*j) Thus, while deep learning has improved melody generation, existing methods either **lack structure (ML-***

based) or lack creativity (rule-based), highlighting the need for a **hybrid approach.**

V. PROPOSED SYSTEM

GenMelody is a hybrid, AI-driven system designed for automatic and multilingual song generation. It addresses several key limitations observed in existing systems such as MusicLM, including challenges related to data quality, high computational requirements, limited multilingual support, and inefficiencies in real-time processing. GenMelody integrates a ReactJS-based frontend, a FastAPI backend, and a robust core engine built around the GPT-2 Medium model to generate high-quality lyrics and melodies effectively.

The use of GPT-2 Medium, replacing the earlier DistilGPT-2 implementation, brings significant improvements in contextual understanding and lyrical coherence. Its deeper architecture allows for more expressive and emotionally resonant lyric generation while remaining computationally viable for real-time usage. Additionally, the system supports multilingual capabilities by automatically translating generated English lyrics into Telugu and Hindi using the Google Translate API, thus broadening its usability across diverse linguistic backgrounds.

The system architecture is lightweight and modular, designed to optimize scalability and real-time performance. This makes GenMelody suitable for integration into various applications, including web-based composition tools and mobile platforms. The explainable and modular workflow also enhances system transparency, making it easier to debug, interpret results, and perform targeted upgrades or optimizations.

The application is composed of several key components. The frontend, built with ReactJS, provides a user-friendly interface for entering text phrases and viewing the resulting multilingual lyrics and audio. The FastAPI backend serves as a communication bridge between the frontend and the core generation engine, handling all routing and data flow. At the core, the generator module utilizes a fine-tuned GPT-2 Medium model to create lyrical content. This output is then processed by the translation module to support Hindi and Telugu. The melody generation process leverages sentiment analysis and the Music21 library to compose musical sequences that align with the mood and rhythm of the lyrics. Finally, the voice synthesis module converts lyrics into audio using gTTS and overlays them with melodies using FluidSynth and Pydub, resulting in complete song tracks.

GenMelody efficiently produces coherent, musically-aligned lyrical content in multiple languages, supporting seamless transitions between English, Telugu, and Hindi. Its combination of modern AI techniques, lightweight infrastructure, and real-time responsiveness makes it a

powerful tool for both amateur users and professional musicians seeking creative automation.

VI. METHODOLOGY

GenMelody’s development builds upon a variety of melody generation strategies, combining their strengths into a unified framework.

A. Rule-Based Generation

This component employs predefined musical grammar, such as scale and chord progression rules, to ensure the theoretical correctness of generated melodies. While structurally sound, rule-based melodies may lack variation and emotional depth.

B. Statistical Models

Markov Chains are used to model probabilistic transitions between notes, offering a basic level of variation. Hidden Markov Models (HMMs) enhance this approach by introducing hidden states that capture underlying musical intentions, resulting in more diverse melodic patterns.

C. Deep Learning-Based Models

RNNs and LSTMs learn temporal dependencies in musical sequences, contributing to smoother and more coherent melodies. Transformers extend this capability by modeling long-range relationships using attention mechanisms. Additionally, Variational Autoencoders (VAEs) are explored to introduce controlled randomness and generate diverse, yet structured, melodic outputs.

D. Hybrid Approaches

The final system integrates rule-based constraints with AI-generated melodies. This fusion enables GenMelody to enforce harmonic correctness while allowing for creative improvisation, achieving a balanced and expressive musical composition process.

TABLE I. COMPARISON OF MELODY GENERATION TECHNIQUES

S.N O	Different Methods		
	Method	Advantages	Limitations
1	Rule-Based	Ensure harmonic Correctness	Lacks creativity
2	Markov Models	Probabilistic structure	Poor long-term dependency
3	LSTMS	Captures sequential patterns	Struggles with long melodies
4	Transformers	Strong long-term structure	High computational cost
5	Hybrid Models	Balances structure & creativity	Complex implementation

Internal Working of the Proposed System

GenMelody integrates several technologies for seamless song creation in multiple languages, combining lyric generation, music synthesis, translation, and voice generation to produce complete multilingual songs. The system consists of three primary functions:

1. Generative Engine: GPT-2 Medium

Overview:

GPT-2 Medium is a larger version of the GPT-2 model, offering more robust language generation capabilities compared to smaller variants like DistilGPT-2. It has a greater capacity for generating more nuanced and contextually rich lyrics, making it ideal for lyric generation in *GenMelody*.

Lyric Generation: Upon receiving an input prompt, the system tokenizes the phrase and generates contextually relevant, coherent lyrics, ensuring appropriate rhyme schemes and rhythmic structures. The greater capacity of GPT-2 Medium allows for more varied and expressive lyrical generation.

Postprocessing: Generated lyrics undergo cleanup to ensure proper formatting, grammar, and coherence before being stored in the central database.

2. Translation and Voice Generation

Googletrans (Google Translate API): Googletrans is used to translate English lyrics into Telugu and Hindi. The system carefully handles the translation process to preserve the poetic and emotional essence of the lyrics, applying post-processing rules to avoid issues with literal translations.

gTTS (Google Text-to-Speech): After translation, gTTS converts the text into audio, adjusting pitch and rate to simulate a singing voice. The resulting MP3 files are integrated with MIDI melodies to create a complete multilingual song.

3. Music Synthesis

Fluidsynth: Fluidsynth is used to convert MIDI events into audio, with musical instruments selected based on the song’s mood. The generated MIDI file is then converted into raw audio.

IndicNLP Library: The IndicNLP library ensures phonetic accuracy for Telugu and Hindi lyrics by splitting syllables and aligning them with the melody.

Music21: Music21 is employed for musical analysis and composition, converting lyrical syllables into musical notes and incorporating musical theory features such as key signatures, chord progressions, and dynamics.

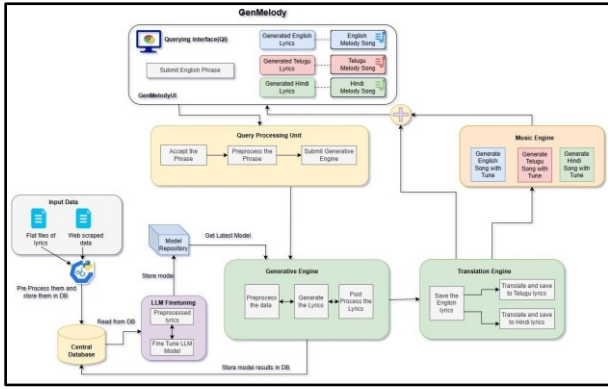


Fig. 1. **Proposed Architecture of GenMelody:** consists of the User Interface, Query Processing Unit, Generate-Translate-Music Engines, and Custom Fine-Tuning Support for enhanced music generation.

VII. RESULTS AND DISCUSSION

Experiments were conducted using a dataset of **symbolic music sequences** to evaluate different models. Explored different CPU based GPT models aka distillgpt2, gpt2-medium in Personal Computer lab setup, the results are here as follows:

Prompt 1: walking in the sunshine with my girl Distillgpt2:

walking in the sunshine with my girl.
 I was in a car with her and I was driving to the park. I saw park. I saw the first thing I noticed was the size of the screen. It was a little too big for me big for me.
 I'm not going to be able to do that. to do that.
 I'm not going to do it. I'm walking in the sunshine with my girl.
 I was in a car with her and I was driving to the park. I saw park. I saw the first thing I noticed was the size of the screen. It was a little too big for me big for me.
 I'm not going to be able to do that. to do that.
 I'm not going to do it.

Sample Telugu and Hindi Lyrics

Telugu:

నా అమ్మాయితో సూర్యరశ్మిలో నడవడం.
 నేను ఆమెతో కారులో ఉన్నాను మరియు నేను పార్కుకు డ్రైవింగ్ చేస్తున్నాను. నేను చూశాను
 పార్క్. నేను గమనించిన మొదటి విషయం స్క్రీన్ పరిమాణం. ఇది నాకు కొంచెం పెద్దది
 నాకు పెద్దది.
 నేను అలా చేయలేను.
 అలా చేయడానికి.
 నేను దీన్ని చేయను.

నా అమ్మాయితో సూర్యరశ్మిలో నడవడం.
 నేను ఆమెతో కారులో ఉన్నాను మరియు నేను పార్కుకు డ్రైవింగ్ చేస్తున్నాను. నేను చూశాను
 పార్క్. నేను గమనించిన మొదటి విషయం స్క్రీన్ పరిమాణం. ఇది నాకు కొంచెం పెద్దది
 నాకు పెద్దది.
 నేను అలా చేయలేను.
 అలా చేయడానికి.
 నేను దీన్ని చేయను

Hindi:

मेरी लड़की के साथ धूप में चलना।
 मैं उसके साथ एक कार में था और मैं पार्क में जा रहा था। मैंने देखा पार्क। मैंने देखा कि पहली चीज जो मैंने देखी थी वह स्क्रीन का आकार था। यह मेरे लिए थोड़ा बड़ा था मेरे लिए बड़ा।
 मैं ऐसा करने में सक्षम नहीं होने जा रहा हूँ।
 वैसे करने के लिए।
 मैं ऐसा नहीं करने जा रहा हूँ मैं हूँ
 मेरी लड़की के साथ धूप में चलना।
 मैं उसके साथ एक कार में था और मैं पार्क में जा रहा था। मैंने देखा पार्क। मैंने देखा कि पहली चीज जो मैंने देखी थी वह स्क्रीन का आकार था। यह मेरे लिए थोड़ा बड़ा था मेरे लिए बड़ा।
 मैं ऐसा करने में सक्षम नहीं होने जा रहा हूँ।
 वैसे करने के लिए।
 मैं ऐसा नहीं करने जा रहा हूँ मैं हूँ

Gpt2-medium:

friend
 I was so excited to get out of the house and walk around the neighborhood
 I was excited because I had been planning to go to the beach for a while and I wanted to see the ocean
 But I also wanted a little bit of privacy

The beach was a bit more crowded than I expected
 There were a lot of people walking around and there were lots of cars
 It was pretty crowded
 The beach is a very popular place for people to hang out

People are always walking by and talking to each other
 So I decided to walk by myself
 When I got to my house, I saw a couple of guys walking down the street
 They were talking about something

One of them was talking with a girl
 She was wearing a bikini
 He was walking with her
 And he was looking at her boobs

That was the first time I realized that I wasn't alone

*This was my first experience with naked people
My friend was also naked
We were walking along the sidewalk*

*As we walked along, we saw two guys talking
A guy was saying something to another guy
Then the guy next to me was asking the other guy what he wanted
After a few minutes, the two of us started talking again*

*At this point, my friend started to feel uncomfortable
What are you talking too? I asked
His response was, I'm talking because you're naked
I said, No, you don't have to be naked to talk to someone*

*He said that he didn't want to hear it
Later that night, he told me that the next day he would be walking home alone and he wouldn't be able to stop talking and that it was his first naked conversation
After that, it became a regular thing
Every time we went to a beach, there would always be a bunch of naked guys*

Sometimes they would walk down a street and they'd be talking, and sometimes they wouldn't

Prompt 2: rain makes me sleepy

Distillgpt2:

rain makes me sleepy.

*I'm not sure if I'm awake, but I can
but I can't wait to see what happens next.
what happens next.*

*The first thing I noticed was that the first time I saw the video, I
the video, I'm not sure if it's true.*

Im

rain makes me sleepy.

*I'm not sure if I'm awake, but I can
but I can't wait to see what happens next.
what happens next.*

*The first thing I noticed was that the first time I saw the video, I
the video, I'm not sure if it's true.*

Gpt2-medium:

*I'm not sure if I'm supposed to be happy or sad
I don't know if it's supposed to be sad or happy
It's just not my thing
But I know I should be*

*And I am
Because I have to
If I didn't, I wouldn't be here
So I do*

*Sometimes I feel like I need to cry
Like I can't
Or maybe I just need it
Maybe I want to feel something*

*Something that's not just a feeling
That's something that I've been missing
A feeling that isn't just something I get
Not just an emotion*

*An emotion that doesn't exist
Just something, something real
The feeling of something in my chest
My heart*

*When I cry, it feels like it hurts
As if my heart is breaking
In a way, that hurts me
For a moment*

*Then I realize that it doesn't
Is
Real
There's nothing I could do about it, but I try*

*Even though I think it might hurt
To feel it in the wrong way
With the right words
At the same time*

Prompt 3: Sun is shiny

Distillgpt2:

Sun is shiny and shiny.

*The first thing I noticed was that the color was very different from the other
from the other side of the road.*

*The next day, the police arrived at the scene and arrested the
and arrested the man who allegedly stole the car from a friend.*

*The man, identified as a man from
a man from the Philippines.*

The Philippines is the only country in the world that has a population of more

Sun is shiny and shiny.

*The first thing I noticed was that the color was very different from the other
from the other side of the road.*

*The next day, the police arrived at the scene and arrested the
and arrested the man who allegedly stole the car from a friend.*

*The man, identified as a man from
a man from the Philippines.*

*The Philippines is the only country in the world that has a
population of more*

Gpt2-medium:

the sun is rising

The sun rises and sets

The sun sets and rises again

It is the same thing

There is no difference

You can't see it

But you can feel it, you know?

The results from the conducted experiments demonstrate that the GPT2-medium model yields superior performance compared to the DistilGPT2 model in the task of lyric generation. This can be attributed to the larger model size and training corpus of GPT2-medium, which enables it to produce more contextually accurate and semantically richer outputs. In contrast, DistilGPT2, being a compressed and distilled version of GPT2, exhibits limitations in generating coherent and engaging content.

All experiments were performed on a personal computer lacking GPU acceleration and equipped with limited RAM, which constrained the ability to utilize more recent and computationally intensive models such as OpenAI's GPT-4, Meta's LLaMA 3, Anthropic's Claude, and models provided by Perplexity AI. These models typically require access through paid APIs and robust hardware configurations. Similarly, deployment of the DeepSeek model was not feasible under the current lab setup due to comparable resource constraints.

Despite these limitations, the observed performance differences highlight the impact of model scale and training data on output quality. It is evident that leveraging state-of-the-art language models, under suitable hardware and licensing conditions, would likely yield results that more closely align with human creativity and linguistic expectations than earlier generations of models such as GPT2.

VIII. CONCLUSION AND FUTURE WORK

The proposed **GenMelody** system advances automatic melody generation by integrating AI-driven techniques with rule-based constraints, achieving a balance between creative output and musical structure. By leveraging the capabilities of GPT-2 Medium, the system successfully supports multilingual lyric and melody generation, demonstrating its potential to produce diverse, expressive, and structurally coherent musical content.

Looking ahead, several directions for future work can further enhance the system's capabilities. One promising area is the integration of reinforcement learning (RL) to enable real-time melody adaptation, allowing the system to respond dynamically to user feedback or performance

contexts. Another important development involves designing symbolic-to-audio transformation models, which can convert symbolic music representations into rich and expressive audio outputs, thereby closing the gap between composition and performance. Additionally, incorporating interactive user control mechanisms will empower users to modify melodies in real time, fostering a more engaging and personalized music creation experience.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to Anil Neerukonda Institute of Technology and Sciences for providing the support and infrastructure essential for this research. I extend my heartfelt thanks to my research guide, Ms. Mallika Thogaru, for her invaluable guidance, continuous encouragement, and insightful feedback throughout the course of this work. I am also thankful to Dr. K. Selvani Deepthi, Head of the Department of Artificial Intelligence and Machine Learning, for her support and for fostering a research-friendly environment. I sincerely appreciate the cooperation and constructive suggestions from the faculty members and peers of the Department of Artificial Intelligence and Machine Learning. The resources and facilities made available by the institute played a vital role in the successful completion of this research.

REFERENCES

- [1] G. F. Welch, M. Biasutti, J. MacRitchie, G. E. McPherson, and E. Himonides, "Editorial: The Impact of Music on Human Development and Well-Being," *Front. Psychol.*, vol. 11, 2020. <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2020.01246/full>
- [2] S. Hallam, "The power of music: Its impact on the intellectual, social and personal development of children and young people," *Int. J. Music Educ.*, vol. 28, pp. 269-289, 2010. https://www.researchgate.net/publication/242560857_The_power_of_music_Its_impact_on_the_intellectual_social_and_personal_development_of_children_and_young_people
- [3] Chen, Yihao & Lerch, Alexander. (2020). Melody-Conditioned Lyrics Generation with SeqGANs. https://www.researchgate.net/publication/348697986_Melody-Conditioned_Lyrics_Generation_with_SeqGANs
- [4] Kashyap, Abhishek & Chauhan, Adarsh & Arya, Bhavishya & Sharma, Ayush. (2023). Lyrics Generator Using RNN. https://www.researchgate.net/publication/369559468_Lyrics_Generator_using_RNN
- [5] Barbieri, Gabriele & Pachet, Francois & Roy, Pierre & Degli Esposti, Mirko. (2012). Markov Constraints for Generating Lyrics with Style. *Frontiers in Artificial*

- Intelligence and Applications. 242. 10.3233/978-1-61499-098-7-115.
<https://www.researchgate.net/publication/236166532> Markov Constraints for Generating Lyrics with Style
- [6] Dhandapani, Aarthi & Nagappan, Ilakiyaselvan & Mandal, Satyaki & Bhadra, Sandipta & Viswanathan, V.. (2023). Lyrics Generation Using LSTM and RNN. 10.1007/978-981-99-1051-9_24.
<https://www.researchgate.net/publication/371484570> Lyrics Generation Using LSTM and RNN
- [7] Vechtomova, Olga, Bahuleyan, Hareesh, Ghabussi, Amirpasha, & John, Vineet. (n.d.). Generating lyrics with variational autoencoder and multi-modal artist embeddings. University of Waterloo, ON, Canada.
<https://arxiv.org/pdf/1812.08318>
- [8] Wang, Lei, Zhao, Ziyi, Liu, Hanwei, Pang, Junwei, Qin, Yi, & Wu, Qidi. (Year). A review of intelligent music generation systems.
<https://arxiv.org/pdf/2211.09124>
- [9] Llewellyn, C. (n.d.). Rule-based algorithms for music generation. Music and Computer Science, Simpson College.
https://www.micsymposium.org/mics_2007/Llewellyn.pdf
- [10] Xu, Yurui. (2023). Music Generator Applying Markov Chain and Lagrange Interpolation. Highlights in Science, Engineering and Technology. 39. 266-273. 10.54097/hset.v39i.6538.
<https://www.researchgate.net/publication/369877578> Music Generator Applying Markov Chain and Lagrange Interpolation
- [11] Van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). WaveNet: A generative model for raw audio. arXiv preprint arXiv:1609.03499
<https://www.researchgate.net/publication/308026508> WaveNet A Generative Model for Raw Audio
- [12] Pal, Abhilash & Saha, Sourav & Ramalingam, Anita. (2020). Musenet : Music Generation using Abstractive and Generative Methods. International Journal of Innovative Technology and Exploring Engineering. 9. 784-788. 10.35940/ijitee.F3580.049620.
<https://www.researchgate.net/publication/363856706> Musenet Music Generation using Abstractive and Generative Methods
- [13] Zhu, Y., Baca, J., Rekabdar, B., & Rawassizadeh, R. (2023). A survey of AI music generation tools and models. arXiv preprint arXiv:2308.12982v1.
<https://arxiv.org/abs/2308.1298>
- [14] Child, R., Gray, S., Radford, A., & Sutskever, I. (2019). Generating long sequences with sparse transformers. arXiv preprint arXiv:1904.10509v1.
<https://arxiv.org/abs/1904.10509>
- [15] Tian, L., & Yang, X. (n.d.). DeepLyrics: GPT2 for lyrics generation with finetuning and prompting techniques (Stanford CS224N Custom Project). Stanford University.
<https://cs224n.stanford.edu/>
- [16] Agostinelli, A., Denk, T. I., Borsos, Z., Engel, J., Verzetti, M., Caillon, A., Huang, Q., Jansen, A., Roberts, A., Tagliasacchi, M., Sharifi, M., Zeghidour, N., & Frank, C. (2023). MusicLM: Generating music from text. arXiv preprint arXiv:2301.11325v1.
<https://arxiv.org/abs/2301.11325>
- [17] Huang, R., Huang, J., Yang, D., Ren, Y., Liu, L., Li, M., Ye, Z., Liu, J., Yin, X., & Zhao, Z. (2023). Make-An-Audio: Text-to-audio generation with prompt-enhanced diffusion models. Stanford CS224N: Natural Language Processing with Deep Learning. Stanford University. Retrieved from
<https://cs224n.stanford.edu/>
- [18] Jiao, X., Yin, Y., Shang, L., Jiang, X., Chen, X., Li, L., Wang, F., & Liu, Q. (2020). TinyBERT: Distilling BERT for natural language understanding. arXiv preprint arXiv:1909.10351v5.
<https://arxiv.org/abs/1909.10351>
- [19] Li, S.; Sung, Y. MRBERT: Pre-Training of Melody and Rhythm for Automatic Music Generation. Mathematics 2023, 11, 798.
<https://www.researchgate.net/publication/368319102> MRBERT Pre-Training of Melody and Rhythm for Automatic Music Generation
- [20] Briot, JP., Pachet, F. Deep learning for music generation: challenges and directions. *Neural Comput & Applic* 32, 981–993 (2020).
<https://doi.org/10.1007/s00521-018-3813-6>