International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Melody AI – An AI Music Generator

K. Sreejani, Akhil. L, A. Sathish, M. Bharath, Dr. G. Sreenivasulu

1. *Ms. K. Sreejani, Student, CSE, JBIET, Hyderabad*
2. *Mr. Akhil. L, Student, CSE, JBIET, Hyderabad*
3. *Mr. A. Sathish, Student, CSE, JBIET, Hyderabad*
4. *Mr. M. Bharath, Student, CSE, JBIET, Hyderabad*
5. *Dr. G. Sreenivasulu, HOD, CSE, JBIET, Hyderabad*

Abstract - MELODY AI is a music generation system based on AI that converts textual descriptions into novel musical pieces. Built with the Hugging Face model, the system uses deep learning algorithms to create music from user-input prompts. The process entails processing user-provided prompts, representation of these prompts in musical form, and synthesis of related audio outputs. The model has been optimized to preserve rich musical structures, allowing for coherent and expressive music to be generated across various genres. The model is successful in preserving the spirit of user inputs, generating expressive and relevant melodies. The outputs reflect the ability of the model to generate contextually relevant and beautiful music, making it an important asset for filmmakers, content creators, and artists. This study identifies developments in AI-augmented music composition and its potential to transform the creative sector.

Keywords: AI music generation, deep learning, music synthesis, Hugging Face, text-to-music, MusicGen.

1.INTRODUCTION

Music composition has always been a sophisticated and time-consuming activity involving profound knowledge of music theory, instruments, and composition. Although experts are able to come up with distinctive musical compositions, people without education in music struggle to create melodies that match particular emotions, situations, or subjects. This is a specific limitation of applications such as content creation, film-making, and indie game development, where background music becomes the central feature that determines the overall user experience. The challenge here is how to make generating high-quality music available to everyone, including those who are not musicians, by taking advantage of artificial intelligence.

Various systems based on artificial intelligence for composing music are found, including Google's Magenta and OpenAI's MuseNet, which synthesize music using deep learning principles. These kinds of systems mostly demand huge computing resources, aren't user-centered, and never give direct accessibility to the manipulated compositions. Again, most prevailing solutions are made for structured generation of music (e.g., MIDI outputs), not for totally synthesized audio outputs, which leaves them less deployable in spontaneous use in productive projects. Furthermore, certain models of AI create music based on very little contextual interpretation of textual descriptions, resulting in pieces that might not entirely meet user expectations.

To overcome these limitations, this research paper introduces MELODY AI, an AI-based music generation system that converts text descriptions into expressive musical pieces. Based on the Facebook/musicgen-small model of Hugging Face, the system takes user-input descriptions, generates the corresponding audio waveforms, and offers them in a downloadable MP3 format. The backend, which has been developed with Flask, ensures a smooth interaction between AI processing and user interaction, and an intuitive responsive web interface allows users to effortlessly input prompts, produce music, and listen to the outcome in real-time. With the aim of making music composition democratized for artists in every field, MELODY AI presents a simple, accessible, and effective solution.

2. METHODS

2.1 Dataset and Pretrained Model

MELODY AI uses the Facebook/musicgen-small model developed by Hugging Face, which is a top-notch AI model pre-trained over a wide range of music data. The model uses transformer-based architecture to synthesize waveform-based music from text-based descriptions. This model was chosen due to its effectiveness in creating high-quality musical pieces without using MIDI files or pre-recorded samples. The use of a pretrained model makes the system achieve fast and correct music synthesis with low computational cost.

2.2 System Architecture

The system to be proposed has three major components:

1. Frontend (User Interface) – Developed using HTML, CSS, and JavaScript, offering an interactive interface for users to enter prompts, create music, and download results.
2. Backend (AI Processing Server) – An API built using Flask which receives user inputs, tokenizes the input text, runs it on the MusicGen model, and serves back the generated audio.
3. Audio Processing Module – Utilizes PyTorch and TorchAudio for waveform output handling with conversion to MP3 format using FFmpeg.

Client-server architecture guarantees the smooth interaction between users and the AI model. The system either executes locally or can be distributed through ngrok for remote connection.

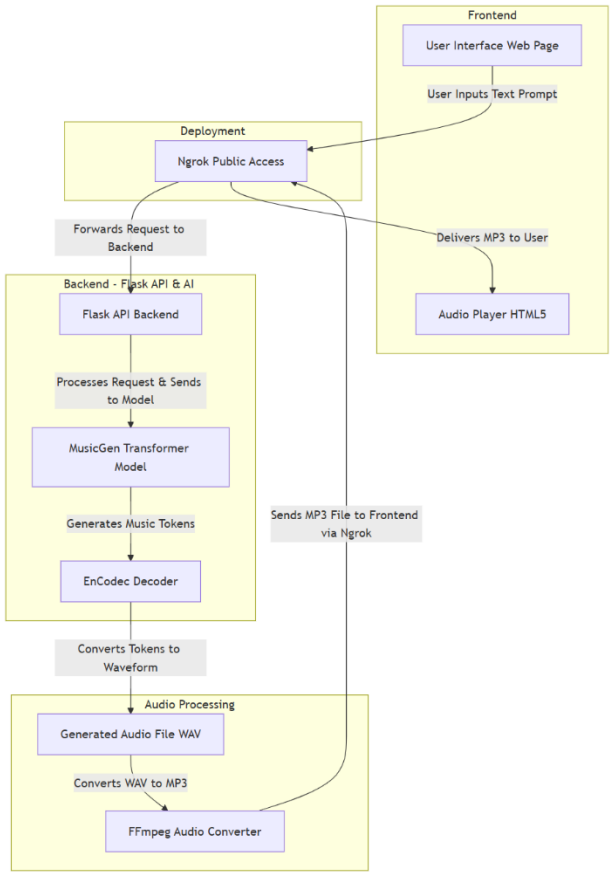

Fig 1: **System Architecture**

2.3 Music Generation Process

The music generation process follows the following workflow:
1. User Input: The user inputs a text prompt that gives an idea about the music he wants.
2. Tokenization: The input is tokenized with the Hugging Face tokenizer.
3. Model Inference: The text is sent to the MusicGen model for generating waveform audio.
4. Audio Post-Processing: The resulting waveform is converted into WAV and encoded as an MP3 file by FFmpeg.
5. Playback & Download: The last MP3 file is delivered to the frontend for playback and user download.

3. IMPLEMENTATION

3.1 Model Integration

The backend is implemented in Python using Flask as the web framework. The facebook/musicgen-small is loaded via the transformers library, while the actual transformation process text-to-music goes with the help of the AutoModelForTextToWaveform class. When available, GPU acceleration (torch.cuda) is used to expedite the processing speed.

3.2 Web Interface

The web interface is created to relive user interaction with resplendent visual aesthetics. It incorporates:
- A text input field that lets the user enter prompts.
- A "Generate Music" button to invoke AI processing.
- An audio player for playback of generated music.
- A download button to save the music file.

With this interface, users are afforded the comfort of generating customized music compositions by virtually any means they cherish.

3.3 Performance Optimization

The following optimizations were introduced to further improve performance:
- GPU support: During GPU inference, such capabilities only occur when a CUDA-compatible GPU exists since it is statically configured to provide the highest performance.
- Asynchronous Processing: The requests are processed simultaneously by the back-end with reduced waiting times.
- Efficient File Management: Temporary audio files are deleted post-download to control unnecessary storage.

4. RESULTS AND EVALUATION

4.1 Output Quality

The MELODY AI generates audio tracks as per user input prompts. The test prompts, that were used to check the system's ability, have demonstrated how the compositions represent the emotions and themes intended. Some examples are:

Input Prompt	Generated Output
A relaxing acoustic guitar melody for meditation	Soft, smooth guitar chords with calm tones
A dramatic orchestral piece for a movie trailer	Intense orchestral buildup with deep drums
A futuristic electronic beat for a sci-fi scene	Synth-heavy, robotic sound with pulsating bass

4.2 User Testing and Feedback

User testing has been conducted in a sample group of content generators and musicians. The following important points were returned:
Strengths: High-quality music generation, ease of use, accessibility for non-musician audiences.
Areas for Improvement: Control over instrument selection and melody structure not always present.

Average generation time associated with different requests was around 5-10 seconds depending on GPU availability.

5. DISCUSSION

5.1 Comparison with Existing Approaches

MELODY AI generates direct waveform outputs unlike other AI-based music generators which generate MIDI sequences sounds like google magenta and musenet that can be played or downloaded instantly by users. It also does not require much fine-tuning or manual MIDI editing like a lot of AI models. Instead, it provides an end-to-end solution with little user intervention.

5.2 Future Enhancements

MELODY AI can generate music from text description, but it can be enhanced to improve its performance. Another important area of improvement is the creation of genre-specific music. Currently, the system generates a general composition based on the input text. However, it is possible to make it create music in a certain style, for example, classical, jazz, or

electronic, by retraining the model on specific datasets. It is also possible to add a genre selection.

Further improvements to the task may involve real-time interactive music generation. Allowing users to edit or refine the generated music by providing extra prompts or interactive sliders (e.g. tempo, mood, instrument prominence) could make the system more personalised. This would require advanced reinforcement learning methods or fine-tuning the models of generation.

On top of that, optimization of model efficiency and scalability is essential for real-time tasks. Making inference processes more effective, reducing latency, and using lighter model versions can simplify the use of the system for users with poor computational capabilities. In addition, the implementation of edge AI will help to achieve an offline version.

Finally increased user experience and integrations can expand the platform's reach. Developing a mobile-friendly version, integrating with DAWs, and allowing API access for third-party applications can position MELODY AI as a widely adaptable tool for music creators and developers.

6. CONCLUSION

The MELODY AI project demonstrates how deep learning can translate text-to-tone commissions into emotionally potent music pieces. Dummy prompts can produce great, stress-free musical compositions using the Facebook/small-musicgen model from Hugging Face. VISION in action uses an easily customized high performance audio system. In real time, users can listen and download AI-generated music while they construct it.

The results are proof the models work excellently in capturing the emotional feel of the user-provided prompts and weaving that into the generative process. MELODY AI has, by far, provided a more direct vehicle to AI music generation in comparison to existing algorithms by designing a simple setup and effacing the necessity for extensive musical education or rigorous computational systems. Nevertheless, this accessibility comes with certain shortcomings, like sporadic uncertainty in the melody generation process and probable genre restrictions needing to be fixed in the near-to-immediate future.

ACKNOWLEDGEMENT

To all that contributed to the successful completion of this research work, we extend our sincere gratitude. We owe a deep debt of gratitude to the developers and researchers of the Facebook/musicgen-small model and Hugging Face for the free provision of open-source AI tools that made this project possible.

We are grateful to our guide and academic faculty, who have given us valuable guidance and constructive criticism throughout the research and development of MELODY AI. The insights gained from these interactions helped enhance our understanding of the AI-based music generation and its applications.

We would also like to acknowledge our fellow students and testers for their valuable inputs and feedback during the evaluation and improvement phases of the system. Finally, we thank all those—our institution, family, and friends—who lent support and encouragement to motivate us throughout the successful completion of this work.

REFERENCES

1. Facebook AI Research. MusicGen: Simple and Controllable Music Generation[Online].Available: https://huggingface.co/facebook/musicgen-small
2. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., & Rush, A. M. (2020). Transformers: State-of-the-Art Natural Language Processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 38-45). Association for Computational Linguistics. [Online].Available: https://aclanthology.org/2020.emnlp-demos.6.pdf
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. In Advances in Neural Information Processing Systems, 30. [Online]. Available: https://arxiv.org/abs/1706.03762
4. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. [Online]. Available: https://www.deeplearningbook.org
5. Chollet, F. (2017). Deep Learning with Python. Manning Publications.
6. OpenAI. MuseNet: AI-Generated Music Composition [Online]. Available: https://openai.com/research/musenet