Melody AI – An AI Music Generator

Keetha Ajay Kumar, Jalli Srujan Kumar, Gotte Pavani, Kalva Sai Abhinav, Elijah Francis

1. *Mr. Keetha Ajay Kumar, Student, CSE, JBREC, Hyderabad*
2. *Mr. Jalli Srujan Kumar, Student, CSE, JBREC, Hyderabad*
3. *Ms. Gotte Pavani, Student, CSE, JBREC, Hyderabad*
4. *Mr. Kalva Sai Abhinav, Student, CSE, JBREC, Hyderabad*
5. *Mr. Elijah Francis , Assistant Professor, CSE, JBREC, Hyderabad*

---***---

Abstract - MELODY AI is an innovative system designed to generate music using artificial intelligence by transforming textual descriptions into original compositions. Leveraging the MusicGen model from Hugging Face, the system employs advanced deep learning techniques to interpret user input and produce corresponding audio tracks. The workflow includes interpreting the user's prompt, translating it into a musical representation, and synthesizing it into audio form. The AI model has been fine-tuned to maintain musical richness and structural integrity, enabling it to generate expressive, genre-diverse compositions. The system effectively captures the essence of user prompts, delivering melodically coherent and emotionally resonant music. Its ability to generate tailored and aesthetically pleasing music makes it a valuable tool for creators in film, digital media, and the arts. This project highlights the growing role of AI in reshaping music composition and enhancing creative expression.

Keywords: AI-generated music, deep learning, text-to-audio synthesis, Hugging Face, MusicGen, creative AI tools.

1.INTRODUCTION

Creating music has traditionally required a deep understanding of musical concepts, theory, and instrumentation, often demanding years of practice and experience. While trained musicians can produce intricate and emotive compositions, individuals without formal music education often struggle to craft melodies that convey specific moods, themes, or narratives. This limitation presents a challenge in areas like content creation, filmmaking, and indie game development, where music plays a pivotal role in shaping the audience's emotional experience. Bridging this gap requires a solution that empowers non-musicians to create high-quality, emotionally resonant music with ease.

In recent years, artificial intelligence has opened new possibilities in the field of music generation. AI-based tools such as Google’s Magenta and OpenAI’s MuseNet utilize deep learning to compose music. However, many of these tools are resource-intensive, lack intuitive interfaces, and often produce outputs in structured formats like MIDI, which require further processing before they can be used in real-world applications. Additionally, these models may not fully interpret or align with the emotional or contextual cues embedded in user prompts, leading to results that fall short of expectations.

To address these challenges, this paper introduces MELODY AI, an intelligent system capable of generating music directly from textual descriptions. Built using the facebook/musicgen-small model available on Hugging Face, the system interprets user-provided prompts and synthesizes them into downloadable MP3 audio tracks. The backend, developed using Flask, manages communication between the AI model and the user interface. Users interact through a clean, responsive web interface that allows them to enter prompts, generate music, and instantly preview the results. Designed to make music creation more accessible, MELODY AI offers a practical, user-friendly platform for artists, creators, and enthusiasts across all domains.

2. METHODS

2.1 Dataset and Pretrained Model

MELODY AI is powered by the facebook/musicgen-small model, a pre-trained AI music generation system provided by Hugging Face. This advanced model is built on a transformer architecture and is capable of converting natural language text into waveform-based music. It has been trained on a diverse and extensive collection of musical data, enabling it to generate rich and realistic audio compositions. Unlike models that rely on MIDI formats or pre-sampled audio loops, this model directly produces complete music tracks from scratch. Its integration into MELODY AI ensures efficient and accurate music generation while maintaining a low computational footprint, making it suitable for real-time applications and accessible on standard hardware.

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.

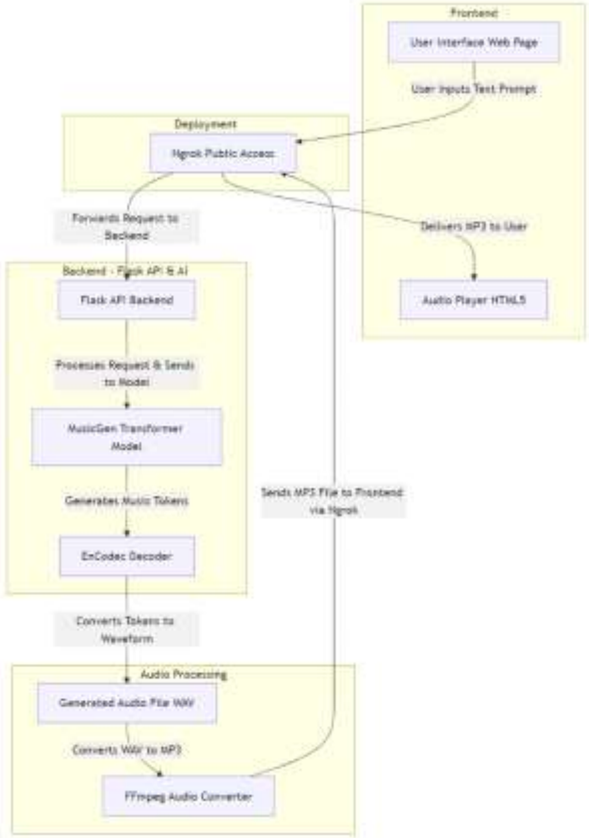

Fig 1: **System Architecture**

2.3 Music Generation Process

The process of generating music in MELODY AI involves a streamlined pipeline consisting of the following steps:
1.Prompt Submission: The user provides a text-based description outlining the desired mood, style, or theme of the music.
2.Text Tokenization: The input text is processed using the Hugging Face tokenizer, converting it into a format suitable for model interpretation.
3.AI Inference: The tokenized prompt is fed into the MusicGen model, which generates raw audio waveforms based on the input description.

4.Audio Conversion: The generated waveform is transformed into a WAV file and subsequently encoded into an MP3 format using FFmpeg for efficient storage and playback.
5.Output Delivery: The final MP3 track is sent to the user interface, allowing users to instantly listen to or download the generated music.

3. IMPLEMENTATION

3.1 Model Integration

The backend of MELODY AI is built using Python, with Flask serving as the primary web framework for handling HTTP requests and responses. The music generation model facebook/musicgen-small is integrated through the Hugging Face transformers library. The core of the text-to-audio conversion process is managed using the AutoModelForTextToWaveform class. To enhance performance, the system checks for the availability of GPU resources (via torch.cuda) and utilizes them when present, significantly reducing inference time.

3.2 Web Interface

A user-friendly and visually appealing web interface has been designed to enhance the overall experience. Key features include:
-A prompt input box for users to type in their music description.
-A “Generate Music” button that triggers the backend model for audio generation.
-An embedded audio player to preview the generated track.
-A download option to save the resulting music file locally.

This simple and responsive design allows users to effortlessly create music tailored to their preferences.

3.3 Performance Optimization

To improve efficiency and ensure smooth operation, several performance enhancements have been implemented:
-GPU Utilization: If a CUDA-compatible GPU is detected, the system automatically switches to GPU-based inference for faster processing.
-Asynchronous Handling: Backend tasks are executed asynchronously to minimize delays and handle multiple user requests concurrently.
-Smart File Management: Temporary audio files are automatically deleted after download to conserve storage and maintain system cleanliness.

4. RESULTS AND EVALUATION

4.1 Output Quality

MELODY AI successfully produces music tracks that align with the emotional tone and style described in user prompts. Various test cases were used to evaluate the system’s responsiveness to different themes. The generated outputs effectively reflected the desired mood and musical characteristics. Below are a few sample prompts and corresponding outcomes:

Input Prompt	Generated Output
A relaxing acoustic guitar melody for meditation	Gentle, soothing guitar strums with mellow, calming ambiance
A dramatic orchestral piece for a movie trailer	Powerful orchestral layers with cinematic intensity and deep drums
A futuristic electronic beat for a sci-fi scene	Robotic synths with rhythmic, pulsing bass and digital textures

These examples highlight the system's ability to generate expressive and contextually appropriate audio content

4.2 User Testing and Feedback

To evaluate user satisfaction and real-world usability, the system was tested by a group of digital content creators and musicians. Feedback revealed the following insights:
-Strengths: The system was praised for its high-quality audio generation, intuitive user interface, and accessibility for individuals without a music background.
-Suggestions for Improvement: Users noted that finer control over specific musical elements—such as instrument choice and compositional structure—could enhance the experience further.
-On average, music generation took approximately 5–10 seconds per request, with faster times observed when GPU acceleration was available.

5. DISCUSSION

5.1 Comparison with Existing Approaches

Unlike many existing AI-based music generation tools such as Google's Magenta or OpenAI's MuseNet, which primarily produce MIDI sequences requiring further editing or conversion, MELODY AI directly generates fully synthesized audio waveforms. This allows users to instantly play or download the final compositions without needing additional software or manual adjustments. MELODY AI simplifies the process by offering an end-to-end solution that operates with minimal user input, eliminating the complexity typically involved in MIDI-based systems.

5.2 Future Enhancements

While MELODY AI is already capable of translating text prompts into expressive musical pieces, several enhancements could further refine its capabilities. One major area of development involves introducing genre-specific generation. Currently, the model produces general music interpretations, but with retraining on curated genre-specific datasets (e.g., classical, jazz, electronic), it could generate compositions in targeted musical styles. Adding a genre selection feature to the interface would further personalize the user experience.

Another significant enhancement involves real-time interactive control. By allowing users to modify or guide the music generation process through additional prompts or adjustable parameters such as tempo, mood, or instrument balance, the system could offer a more customized and dynamic output. Implementing such interactivity would require advanced model tuning, possibly involving reinforcement learning or user-feedback-driven optimization.

Moreover, performance scalability and optimization will be critical for broader adoption. Reducing processing delays, minimizing hardware requirements, and deploying lighter model versions could ensure smooth operation even on devices with limited computing power. Incorporating Edge AI capabilities may also enable an offline version of the system for mobile or low-resource environments.

Lastly, improving user experience and integration options will help expand MELODY AI's ecosystem. Developing a responsive mobile interface, offering plugin support for digital audio workstations (DAWs), and enabling API access for third-party developers would significantly enhance its utility and reach, transforming it into a versatile tool for music production and creative innovation.

6. CONCLUSION

MELODY AI showcases the potential of deep learning technologies in converting textual descriptions into meaningful and emotionally rich musical compositions. By leveraging the facebook/musicgen-small model from Hugging Face, the system enables users—even those without any background in music—to generate high-quality audio tracks with minimal effort. The platform provides a responsive and user-friendly experience where music can be generated, previewed, and downloaded in real time.

The outcomes of the system affirm its effectiveness in understanding and translating user intent into fitting musical outputs. Compared to traditional approaches, MELODY AI simplifies the entire process by eliminating the need for complex MIDI editing or musical theory knowledge, making it accessible to a broader audience. Despite its success, the system still faces limitations such as occasional unpredictability in generated melodies and lack of robust genre-specific customization. These issues present opportunities for future development.

In essence, MELODY AI bridges the gap between creativity and technology, offering a practical solution for AI-powered music generation and setting the stage for more advanced, interactive, and personalized audio creation tools in the future.

ACKNOWLEDGEMENT

We would like to express our heartfelt thanks to everyone who played a role in the successful development and completion of this project. Our deepest appreciation goes to the researchers and developers behind the facebook/musicgen-small model and Hugging Face, whose commitment to open-source AI innovation made this work feasible.

We are especially thankful to our faculty advisor and academic mentors for their continuous support, insightful suggestions, and critical feedback throughout the course of our research. Their guidance was instrumental in shaping our understanding of AI-driven music generation and refining the MELODY AI system.

We also extend our gratitude to our peers and test users, whose observations and feedback during testing phases were vital in improving the quality and performance of the system. Lastly, we are indebted to our institution, families, and friends for their constant support, encouragement, and motivation, which kept us focused and committed to completing this work successfully.

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