Jiaxing YU, Songruoyao WU, Guanting LU, Zijin LI, Li ZHOU, Kejun ZHANG, 2024. Suno: potential, prospects, and trends. *Frontiers of Information Technology & Electronic Engineering*, 25(7):1025-1030. https://doi.org/10.1631/FITEE.2400299

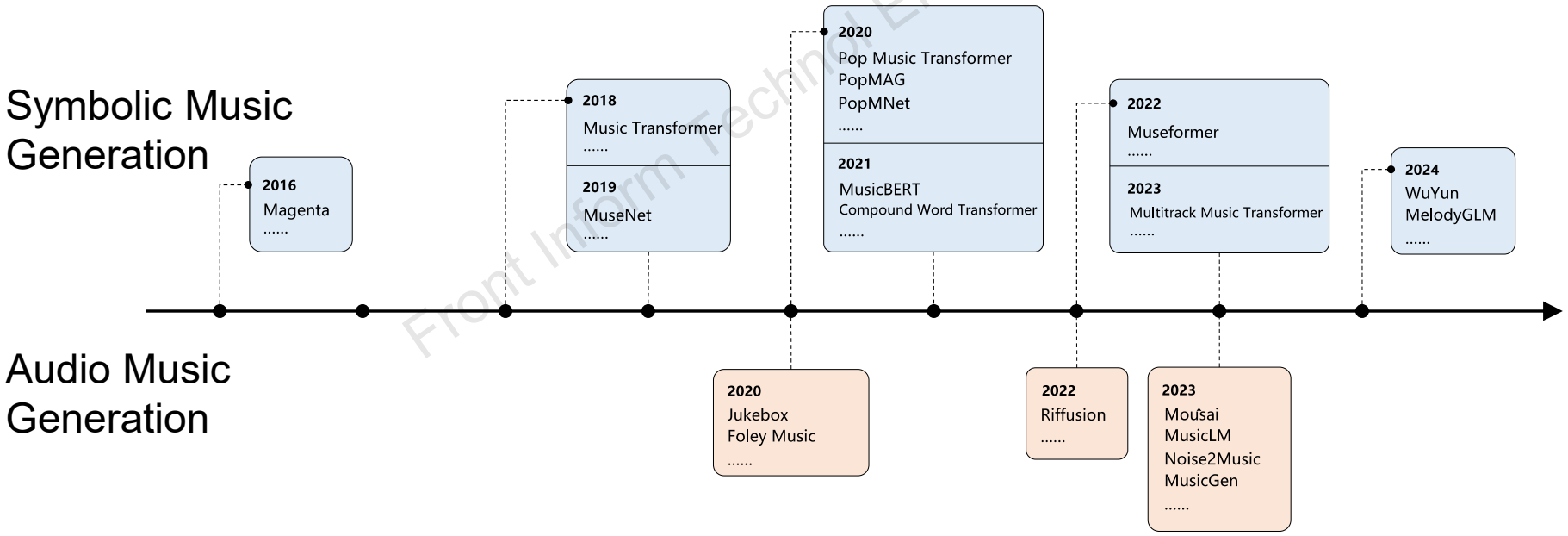
Suno: potential, prospects, and trends

Key words: Music and AI; Music generation; AI music generation platform; Suno

Corresponding author: Kejun ZHANG
E-mail: zhangkejun@zju.edu.cn
ORCID: https://orcid.org/0000-0003-4592-1818

Background

- In recent years, the advancement of neural networks has significantly contributed to the progress of AI music generation, which contains two main categories: symbolic music generation and audio music generation.



Timeline of AI music generation technology

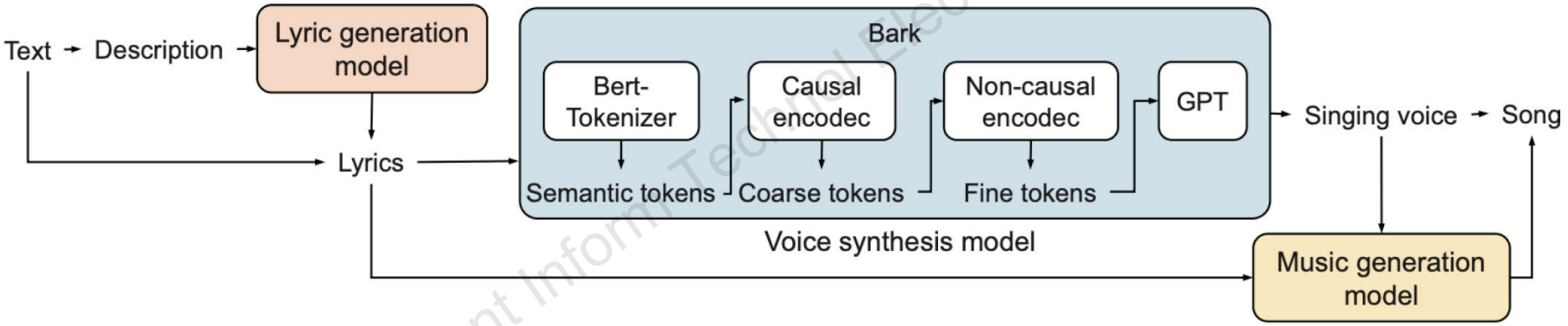
Background

- Symbolic music generation considers music pieces as sequences of text-like information, converting musical elements, such as pitch, position, duration, and velocity, into symbols (MIDI, piano roll, etc.) for processing, modeling, and recreation. It achieves high-quality music generation at a low computing cost, and exhibits characteristics of efficient information encoding, understanding of complex musical structures, and editability.

- Audio music generation refers to a series of techniques that use AI algorithms and models to process, analyze, and generate audio music content, such as waveform and spectrogram. It features richness of information, authenticity of expression, and diversity of generation.

General Technical Framework

- The general technical framework of AI music generation primarily consists of three modules: a lyric generation model, a voice synthesis model, and a music generation model.



General framework of AI music generation, including the lyric generation model, the voice synthesis model, and the music generation model

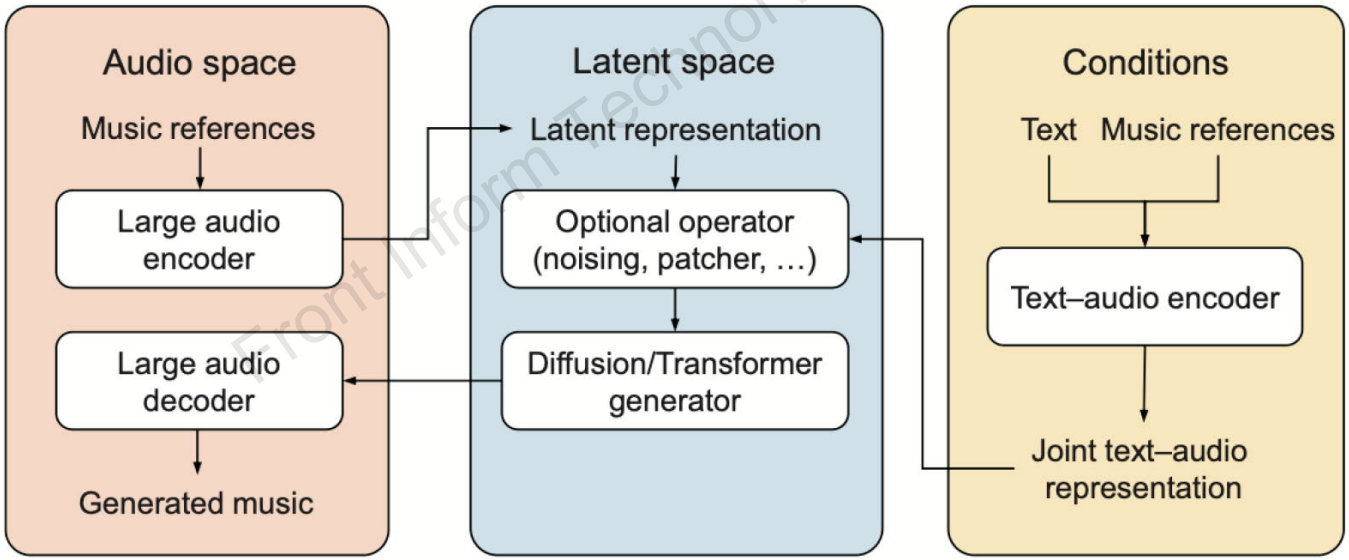
General Technical Framework

- Lyric generation model: A lyric generation model employs advanced natural language processing technologies, enabling a deep understanding of complex user instructions and emotional expressions. This module is featured by its high flexibility and innovation, which allow the model to quickly generate lyrics that match the text descriptions provided by users.

- Voice synthesis model: A voice synthesis model transforms lyrics into natural singing voices. It not only preserves the meaning and emotions of the lyrics but also produces singing voices that closely match human performances.

General Technical Framework

- Music generation model: It takes music references (such as singing voices) as the input and generates corresponding music as the output. Additionally, conditions, such as text descriptions and music references, can ensure that the generated music conforms to specific preferences.



Overall architecture of music generation models that take music references (such as singing voices) as the input and output the generated music

Potential and Prospects

- Suno shows great potential in AI music generation. Its main advantages include:

- Integrated lyric–melody–song creation: Existing AI music generation technologies typically focus on specific aspects of song creation and do not encompass the entire creative process. In contrast, Suno builds a multi-modal music composition platform that integrates the latest advancements in lyric generation, voice synthesis, and music generation.
- Simplified and diverse music composition: Suno makes simplified and diverse music creation possible. Unlike existing music generation technologies that require complex text inputs with musical knowledge, Suno stands out by providing music composition solutions based on simple text descriptions.
- Cross-cultural communication and integration: Suno effectively promotes cross-cultural music communication and integration. As one of the most widely used AI music generation platforms, it breaks down geographical and cultural boundaries and facilitates the blending of musical elements from different cultures, opening up new paths for the development of global music culture.

Limitations

- Suno has demonstrated strong capabilities in understanding and generation of music, but it still has several limitations:
  - Long-term coherence: Suno may face issues establishing coherence between different sections of longer songs. This highlights the challenge of long-term dependencies encountered by the Transformer architecture.

(a)

Example of Suno's limitations: long-term coherence—Suno may face problems with continuity between different parts of longer songs

Limitations

- Fine-grained creation: Suno excels in basic song generation but still struggles to meet more detailed creative demands. For instance, when users attempt to mix multiple music styles or set tempos, Suno often fails to do so perfectly. To address this challenge, one potential solution is to implement a conversational generation platform that allows users to edit generated music pieces through dialog.

(b)

Example of Suno's limitations: fine-grained creation—the tempo of the generated song is 125 BPM, which does not match the description

Limitations

- Multilingual support: Suno exhibits variations in its creative capabilities when dealing with different languages. For example, the model shows inadequacies in understanding the semantics and cultural context of Chinese lyrics. This challenge stems from the complexity of the language and the culture, requiring the model to parse literal meanings and capture subtle cultural nuances, which are crucial for song composition.

(c)

Example of Suno's limitations: multilingual support—Suno is not good at dealing with Chinese polyphonic characters; for example, "朝阳" here is sung as "cháo yáng" by Suno, but the correct pronunciation is "zhāo yáng"

Discussions and Conclusions

- Suno's emergence has already sparked discussions about Music and AI. Here, we discuss several future trends in Music and AI:

- Music creation paradigms: With advancements in tools like Suno, Music and AI is moving toward more interactive composition experiences. This new paradigm enables music generation based on user descriptions and refines the creation process through dialog.
- Music cognition: The application of AI music technologies will profoundly impact our cognition of music. The involvement of AI expands our understanding of music. By analyzing a large number of music pieces, AI can discover previously overlooked patterns and trends in music creation, thereby advancing the development of music theories.

Discussions and Conclusions

- Music industry and ecosystem: The further development of emerging AI technologies will revolutionize the music industry and its ecosystem. Many aspects of music will become more efficient, personalized, and diverse. It will lower the threshold for music generation, making it easier for creators to enter the field, thereby increasing the diversity of musical works.
- Music aesthetic evaluation: Suno leads to discussions on the aesthetic value of AI-generated music. Formulating relevant evaluation standards can help us better understand and appreciate them. It may challenge traditional musical aesthetic theories, prompting us to think about how to improve aesthetic value in technological progress and gain a deeper understanding of music as a complex form of cultural and emotional expression.

Jiaxing Yu has been studying for the PhD degree in design at the College of Computer Science and Technology, Zhejiang University. He received the BE degree from the Chu Kochen Honors College at Zhejiang University. His main research interests include natural language processing, music information retrieval, multimedia, and AI music generation.

Songruoyao Wu has been studying for the PhD degree in artificial intelligence at the College of Computer Science and Technology, Zhejiang University. She received her BE degree from the College of Computer Science and Technology at Zhejiang University. Her main research interests include AI music generation and AI-assisted art creation.

Guanting Lu has been studying for the BE degree in computer science and technology at the Chu Kochen Honors College, Zhejiang University. His main research interests include artificial intelligence and music generation.

Zijin Li is a professor with the Department of Music Artificial Intelligence and Music Information Technology at the Central Conservatory of Music. She also serves as a researcher at CIRMMT and the Creative AI Joint-Lab at QMUL & THU. Her research interests include music acoustics, new instrument design, and the interdisciplinary research paradigm of music and technology.

Li Zhou is an associate professor with the Department of Music at the School of Arts and Communication, China University of Geosciences (Wuhan). She has published over 10 papers in SCI and CSSCI journals and has been granted nearly 20 national invention patents. Her research interests include AI composition, music affective computing, and intelligent music robots.

Kejun Zhang is a professor with the College of Computer Science and Technology, Zhejiang University. He is also director of the Department of Industrial Design and a deputy director of the International Design Institute. He has published more than 40 papers in international journals and conferences. His research interests include artificial intelligence, affective computing, digital humanities, and intelligent art.