Emerging AI Technologies for Music: Towards Controllable, Collaborative, and Creative Systems

Keshav Bhandari K.BHANDARI@QMUL.AC.UK
Queen Mary University of London

Abhinaba Roy ABHINABA_ROY@SUTD.EDU.SG
Singapore University of Technology and Design

Dorien Herremans DORIEN_HERREMANS@SUTD.EDU.SG
Singapore University of Technology and Design

Simon Colton S.COLTON@QMUL.AC.UK
Queen Mary University of London

Editors: D. Herremans, K. Bhandari, A. Roy, S. Colton, M. Barthet

1. Introduction

The field of AI and music has witnessed rapid growth in recent years driving notable success across domains such as music information retrieval (MIR), generative modelling, representation learning, source separation, and transcription among several others. The emergence of large-scale music foundation models (e.g., MusicLM, Suno, Riffusion) and multimodal systems combining modalities such as text, audio and video has further expanded the creative possibilities of AI-driven music. However, despite these technical advances, there remains a significant gap between the capabilities of current AI music systems and their practical adoption by musicians, composers, producers, and educators. At the same time, there has been a growing consensus in the community that recognizes the limitations of fully autonomous generation and the need for interactive, interpretable, controllable, and human-centred approaches. In light of this emerging trend to maintain meaningful human agency over the creative process, the editors of this special issue organized the first workshop on Emerging AI Technologies for Music, as part of the Association for the Advancement of Artificial Intelligence (AAAI) conference in 2026 at Singapore.

A number of publications have acknowledged the need for additional research on human-centric systems to successfully integrate AI within the workflow of musicians, composers, educators, and medical professionals Civit et al. (2022); Dash and Agres (2024); Yu et al. (2023); Sun et al. (2024); Zhao et al. (2025). This direction is bolstered by recent developments including a special track on Human-AI Interaction in Creative Arts and Sciences at IJCNN 2025[1] and the emergence of tools such as Google's Magenta Studio[2], Suno's Song Editor[3] and Sony's Diff-A-Riff[4], which highlight the need for controllable, customizable systems to have humans in the loop and be central to the prominent decision-making points.

1. https://2025.ijcnn.org/authors/special-track-papers-and-exhibitions
2. https://magenta.withgoogle.com/studio/
3. https://suno.com/blog/songeditor
4. https://cslmusicteam.sony.fr/news/diffariff-cocreation-models/

This special issue aims to present a collection of studies at the intersection of AI and music that touch upon topics such as interpretability, controllability and collaboration in the context of such systems.

The included papers represent a variety of themes within music information retrieval (MIR) and generative AI-music applications, which have been grouped by us as follows:

1. **Generative Composition and Performance Tools**

   The papers about generative composition and performance tools focus on AI systems designed to assist or automate the creation of new musical content. For instance, Kaliakatsos-Papakostas et al. (2026) develop a transformer-based framework for automatically generating harmonic accompaniments that stay synchronized with a user's melody. Their framework allows flexible conditioning by allowing users to fix specific chords at any point in the sequence. Similarly, Salcedo and Egozy (2026) presents a real-time music visualizer that uses neural cellular automata to respond to audio energy, emphasizing collaboration by performing inference locally to ensure low-latency responsiveness for live performers. In Ram (2026), the author presents a tool that suggests where to place rests in a melody, offering interpretability by providing analysis artifacts (such as phrase boundaries and tension points) and controllability through steerable cultural presets. "SingingSDS" by Han et al. (2026) is a conversational agent that responds to users through singing, facilitating collaboration in role-play scenarios by allowing users to configure personas and voice profiles for more affective interaction.

2. **Timbre, Voice, and Synthesis Manipulation**

   These studies investigate the "texture" of sound and how AI can modify or transfer specific sonic identities. In the paper "Neural Codec Language Model for Controllable Timbre Transfer", Liu et al. (2026a) introduce a method for zero-shot instrument cloning, allowing a model to adopt the sound of any instrument from a short audio clip. Users can control the final output's sound by providing a short 1-5 second reference audio segment. "Conditional Vocal Timbral Technique Conversion" by Hsu and Yang (2026) focuses on converting vocal styles (like whispers or screams) while preserving speaker identity, using dual-attribute modulation. Their framework gives users precise controllability over both timbre and prosody. Hebbar and McFee (2026) present a study titled "Investigating Timbre Representations in CLAP Across Modalities via Perturbations" which explores mappings from high-dimensional embeddings to psychoacoustic features to improve the interpretability of timbre and how it can be manipulated in retrieval tasks.

3. **Music Theory and Cognitive Perception**

   This group of papers examines how AI perceives musical structures and whether it can understand music in the way humans do. For example, Sadek and Bakarji (2026) demonstrate that unsupervised models can spontaneously organize musical data into a geometric "circle of fifths", mirroring traditional music theory. Carone et al. (2026) through their paper "LLMs can read music, but struggle to hear it" benchmark the "hearing" abilities of large language models, finding they are excellent at reading scores (MIDI) but poor at identifying rhythms or chords from raw audio.

4. **Cultural Heritage and Specialized Data**

   These papers address the use of AI in specific cultural contexts or for specialized, non-mainstream musical data. Within this topic, Sudarshan et al. (2026) built a system to identify rhythmic motifs in South Asian dance. This system serves as a collaborative pedagogical tool to help human instructors document and teach traditional beat structures. Hromadka et al. (2026) in their paper "A Novel Diffusion Model Based Approach for Sleep Therapeutic Music" trained a model to generate personalized audio for insomnia treatment and mental health therapy.

5. **Copyright, AI Ethics and Future Directions**

   This group of papers examine the sociotechnical and legal infrastructures of music AI to ensure creative agency is protected and human-centric values are built into future systems. The paper "Postscript on the Musics of Control" by Chen (2026) analyzes how "control" is marketized in the "gear economy" and calls for ethical, design-oriented interventions that redirect creative agency back to human musicians. "TS-RaMIA" by Liu et al. (2026b) propose an auditing tool for artists to detect if their copyrighted symbolic music (such as MIDI files) was used to train a model without permission. Finally, "Prevailing Research Areas for Music AI" is a survey paper by Wei et al. (2026) that identifies future frontiers of AI music, highlighting the need for better interpretability and controllability in foundation models and more effective human-AI interaction.

We hope the readers find this special issue insightful and engaging. With the rapid evolution of AI technologies for music, we hope the papers here inspire continued exploration of human–AI collaboration, fostering creative systems that empower musicians and deepen the expressive possibilities of music-making. We extend our sincere gratitude to the authors, reviewers, and the entire editorial team for their invaluable contributions and support.

Acknowledgments

This workshop is supported by UKRI and EPSRC under grant EP/S022694/1, as well as SUTD's Kickstart Initiative under grant number SKI 2021 04 06 and MOE under grant number MOE-T2EP20124-0014.

References

Brandon James Carone, Iran R. Roman, and Pablo Ripollés. Llms can read music, but struggle to hear it. an evaluation of core music perception tasks. In *Proceedings of the 1st International Workshop on Emerging AI Technologies for Music (EAIM 2026) at AAAI-26*, pages 1–27, 2026.

Yinuo Chen. Postscript on the musics of control. In *Proceedings of the 1st International Workshop on Emerging AI Technologies for Music (EAIM 2026) at AAAI-26*, pages 1–16, 2026.

Miguel Civit, Javier Civit-Masot, Francisco Cuadrado, and Maria J. Escalona. A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. *Expert Systems with Applications*, 209:118190, 2022. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2022.118190. URL https://www.sciencedirect.com/science/article/pii/S0957417422013537.

Adyasha Dash and Kathleen Agres. Ai-based affective music generation systems: A review of methods and challenges. *ACM Comput. Surv.*, 56(11), July 2024. ISSN 0360-0300. doi: 10.1145/3672554. URL https://doi.org/10.1145/3672554.

Jionghao Han, Jiatong Shi, Masao Someki, Yuxun Tang, Lan Liu, Yiwen Zhao, Wenhao Feng, and Shinji Watanabe. Singingsds: A singing-capable spoken dialogue system for conversational roleplay applications. In *Proceedings of the 1st International Workshop on Emerging AI Technologies for Music (EAIM 2026) at AAAI-26*, pages 1–21, 2026.

Devyani Hebbar and Brian McFee. Investigating timbre representations in clap across modalities via perturbations. In *Proceedings of the 1st International Workshop on Emerging AI Technologies for Music (EAIM 2026) at AAAI-26*, pages 1–12, 2026.

Timo Hromadka, Kevin Monteiro, and Sam Nallaperuma. A novel diffusion model based approach for sleep therapeutic music generation. In *Proceedings of the 1st International Workshop on Emerging AI Technologies for Music (EAIM 2026) at AAAI-26*, pages 1–15, 2026.

Ting-Chao Hsu and Yi-Hsuan Yang. Conditional vocal timbral technique conversion via embedding-guided dual attribute modulation. In *Proceedings of the 1st International Workshop on Emerging AI Technologies for Music (EAIM 2026) at AAAI-26*, pages 1–14, 2026.

Maximos Kaliakatsos-Papakostas, Dimos Makris, Konstantinos Soiledis, and Konstantinos-Theodoros Tsamis. Encoder-only transformers for melodic harmonization: Representation emergence and inference strategies. In *Proceedings of the 1st International Workshop on Emerging AI Technologies for Music (EAIM 2026) at AAAI-26*, pages 1–11, 2026.

Sheldon Liu, Tianyu Liu, Deepak Dalakoti, Adithya Suresh, Yueying Teng, Xuefeng Liu, Atanu Roy, Randeep Bhatia, Daniel Hatadi, and Prabhjeet Ghuman. Neural codec language model for controllable timbre transfer in music synthesis. In *Proceedings of the 1st International Workshop on Emerging AI Technologies for Music (EAIM 2026) at AAAI-26*, pages 1–13, 2026a.

Yuxuan Liu, Rui Sang, Peihong Zhang, Zhixin Li, Kunyang Zhang, Shengyuan He, Ye Li, Kaiyi Xu, and Shengchen Li. Ts-ramia: Membership inference attacks for symbolic music generation models. In *Proceedings of the 1st International Workshop on Emerging AI Technologies for Music (EAIM 2026) at AAAI-26*, pages 1–15, 2026b.

Gokul Srinath Seetha Ram. Silence as music: Controllable and interpretable ai for strategic silence placement. In *Proceedings of the 1st International Workshop on Emerging AI Technologies for Music (EAIM 2026) at AAAI-26*, pages 1–13, 2026.

Najla Sadek and Joseph Bakarji. The circle of fifths as latent geometry in bach's well-tempered clavier. In *Proceedings of the 1st International Workshop on Emerging AI Technologies for Music (EAIM 2026) at AAAI-26*, pages 1–13, 2026.

Carlos Mariano Salcedo and Eran Egozy. Artificial dancing intelligence: Neural cellular automata for visual performance of music. In *Proceedings of the 1st International Workshop on Emerging AI Technologies for Music (EAIM 2026) at AAAI-26*, pages 1–14, 2026.

Ankitha Sudarshan, Atharva Vikas Jadhav, and Rohini Srihari. Low-resource rhythm learning of south asian beat structures: Machine learning approaches to nattuvangam. In *Proceedings of the 1st International Workshop on Emerging AI Technologies for Music (EAIM 2026) at AAAI-26*, pages 1–17, 2026.

Jingjing Sun, Jingyi Yang, Guyue Zhou, Yucheng Jin, and Jiangtao Gong. Understanding human-ai collaboration in music therapy through co-design with therapists. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703300. doi: 10.1145/3613904.3642764. URL https://doi.org/10.1145/3613904.3642764.

Megan Wei, Mateusz Modrzejewski, Aswin Sivaraman, and Dorien Herremans. Prevailing research areas for music ai in the era of foundation models. In *Proceedings of the 1st International Workshop on Emerging AI Technologies for Music (EAIM 2026) at AAAI-26*, pages 1–23, 2026.

Xiaofei Yu, Ning Ma, Lei Zheng, Licheng Wang, and Kai Wang. Developments and applications of artificial intelligence in music education. *Technologies*, 11(2), 2023. ISSN 2227-7080. doi: 10.3390/technologies11020042. URL https://www.mdpi.com/2227-7080/11/2/42.

Yujia Zhao, Mingzhi Yang, Yujia Lin, Xiaohong Zhang, Feifei Shi, Zongjie Wang, Jianguo Ding, and Huansheng Ning. Ai-enabled text-to-music generation: A comprehensive review of methods, frameworks, and future directions. *Electronics*, 14(6), 2025. ISSN 2079-9292. doi: 10.3390/electronics14061197. URL https://www.mdpi.com/2079-9292/14/6/1197.