sciendo

JAISCR, 2026, Vol. 16, No. 2, pp. 185 – 214
10.2478/jaiscr-2026-0010

ARTIFICIAL INTELLIGENCE IN MUSIC: A BIBLIOMETRIC AND SYSTEMATIC REVIEW OF CREATION, PERFORMANCE, AND EDUCATION

Fei Tong[1], Dongjing Jiang[2], Qingchong Jiao[2], Albina Isufi[3], Flynnwell Jianfei Zhang[2,3,*]

[1]*School of Arts, Ludong University*
Yantai, Shandong, 264025, China

[2]*DeepCox Intelligence,*
Toronto, Ontario, M4S 2A2, Canada

[3]*Ontario Knowledge Intelligence Centre,*
Toronto, Ontario, M5G 2J5, Canada

[*]*E-mail: flynnwell.zhang@okic.ca*

Submitted: 4th July 2025; Accepted: 27th December 2025

Abstract

The integration of artificial intelligence (AI) into the music domain has catalyzed a transformative shift in how music is composed, performed, and taught. This paper introduces and frames the concept of music intelligence and employs bibliometric and systematic review methodologies to comprehensively analyze music intelligence. Music intelligence encompasses the development and application of intelligent systems that not only automate or enhance traditional musical tasks but also foster new modes of creativity, interaction, and pedagogy. Tracing the evolution from early rule-based systems to modern deep learning and multimodal models, we examine how AI is increasingly embedded in musical workflows. We highlight applications ranging from generative composition and expressive performance interpretation to real-time accompaniment and personalized education. By positioning AI as an active collaborator rather than a mere tool, this study underscores the need for collaborative efforts among computer scientists, musicians, educators, and cognitive scientists to fully realize the potential of intelligent music systems. Our bibliometric analysis indicates an annual growth rate of 14.92%, with China, the US, and the UK contributing 52.9% of global research output. The findings reveal a rapidly expanding interdisciplinary field characterized by increasing international collaboration, methodological diversification, and a growing focus on human-AI co-creativity. However, persistent gaps remain in cultural inclusivity, interpretability, and ethical governance.

Keywords: bibliometric review, music intelligence, aritificial intelligence, music performance, music education

1 Introduction

1.1 Background

In recent years, the rapid advancement of artificial intelligence (AI) has triggered a profound transformation in the music domain, reshaping how music is created, performed, and taught. This emerging field – referred to in this work as *music intelligence* (MI) – represents a new paradigm at the intersection of AI and musical practice. It encompasses developing and applying intelligent systems that not only automate or augment traditional musical tasks but also enable novel modes of creativity, interaction, and learning. From early rule-based composition algorithms [1, 2, 3] to today's advanced deep learning models [4, 5, 6, 7, 8] and multimodal architectures [9, 10, 11], AI has progressively become embedded in musical workflows. As both a technological construct and a research frontier, music intelligence invites interdisciplinary collaboration among computer scientists, musicians, educators, and cognitive scientists, and captures this growing integration by framing AI not merely as a tool, but as an active collaborator in composition [12, 13], performance interpretation [14, 15], real-time accompaniment [16, 17], personalized music education [18, 19], etc.

1.2 Motivation

Music intelligence (MI) challenges traditional boundaries of authorship and pedagogy while opening new opportunities for democratizing music creation and learning through intelligent systems. Research efforts on MI often target narrowly defined tasks, such as melody generation [20, 21, 22], accompaniment prediction [13, 23, 24], or automatic feedback in music learning [25, 26].

[27] systematically examined AI-based music composition by comparing computational models with human creative processes. The study analyzes how deep learning approaches handle fundamental musical elements like melody, harmony, and structure, while assessing their ability to replicate human-like creativity. The authors organized existing methods into a clear taxonomy based on their musical capabilities, providing insights into how current models address music's inherent complexity. However, the review is limited by its primarily theoretical focus, lacking empirical comparisons of model outputs, and does not fully address the subjective nature of musical creativity evaluation. In [28], a comprehensive survey of computational intelligence techniques applied to algorithmic music composition has been made. It systematically examined existing approaches through the lens of three fundamental musical elements: musical form, melody, and accompaniment, and offered a detailed analysis of how evolutionary algorithms and neural networks have been specifically adapted for music generation tasks, highlighting their respective strengths in handling different aspects of composition. The review organizes these techniques into a coherent framework that demonstrates their growing effectiveness in addressing various compositional challenges. Nevertheless, this survey (published in 2016) does not cover recent advances in deep learning and transformer-based models that have significantly advanced the field. Additionally, while it thoroughly documents technical approaches, it provides less insight into the creative and aesthetic outcomes of these computational methods. [29] presented a systematic review of AI-enabled techniques for objective musician performance assessment, analyzing ten key studies that employ deep learning, multimodal analysis, and sensor-based approaches. The survey demonstrates how these methods address traditional subjectivity in evaluating technical proficiency, musical interpretation, and emotional expression, while highlighting their potential applications in education and competitive settings. The weakness of this review is that the authors primarily focused on technological feasibility without sufficient examination of practical implementation challenges in real-world pedagogical contexts. Additionally, while comprehensive in methodological coverage, it lacks critical analysis of cultural biases in algorithmic assessments and provides limited discussion on integrating these tools with traditional evaluation systems. [30] provided a focused review of AI applications in music education, analyzing representative methodologies through interdisciplinary perspectives of pedagogy, psychology, and computer science. While effectively demonstrating AI's potential for enhancing music learning, the review provides limited technical specifications of system implementations and insufficient exploration of practical challenges encountered in actual educational settings, focusing primarily on theoretical frameworks rather than empirical outcomes. [31] systematically examined the didactic potential of generative AI in music education, iden-

tifying nine key implementation areas, including Augmented Reality (AR) and Virtual Reality (VR) applications, intelligent tutoring systems, and composition assistance tools. The review demonstrates AI's capacity to enhance personalized and interactive learning experiences while transforming traditional pedagogical approaches.

Existing studies fail to capture the synergistic interplay between creation, performance, and education, thereby obscuring a holistic understanding of the field's development. Many reviews are either technologically focused with limited discussion on pedagogical or artistic implications, or they concentrate on theoretical frameworks without sufficient empirical or practical grounding [30]. There has not yet been a comprehensive review of music intelligence research that systematically examines the field's three core domains (composition, performance, and education) through both quantitative bibliometric mapping and qualitative critical analysis. These issues point to the need for a more comprehensive understanding of how AI technologies are being applied, developed, and evaluated in musical contexts.

1.3 Contributions

To address this need, this paper provides a comprehensive overview of AI-driven research in music by quantifying publication trends and research focus areas across music creation, performance, and education; categorizing the core AI technologies employed, such as neural networks [32, 33], Transformer models [34], variational autoencoders [35], generative adversarial networks (GANs) [36], diffusion models [37], reinforcement learning [38], and multimodal frameworks [39]; and identifying widely used public models (*e.g.*, MusicLM [40], Jukebox [41], MuseNet [42]) and tools or libraries (*e.g.*, Magenta [43], AIVA [44], OpenAI MuseNet [42], Riffusion [45]). We review and analyze relevant AI research within each of these three domains.

– Bibliometric Analysis of the Literature: We trace publication trends using data from major academic databases including Web of Science (WoS), Institute of Electrical and Electronics Engineers (IEEE) Xplore, and the Association for Computing Machinery (ACM) Digital Library. Our findings reveal increasing interdisciplinary integration, expanding educational applications, and a growing emphasis on human-AI collaborative creativity. We implement bibliometric techniques through Bibliometrix and VOSviewer [46].

– Systematic Review of AI Technologies and Innovations: We examine how state-of-the-art AI methods (*e.g.*, neural networks [32, 33], Transformer [47, 48], GANs [49, 10]) are being deployed in music creation, performance, and education, as shown in Figure 1. These technologies enhance creative productivity, expressiveness, and personalized learning experiences while reshaping pedagogical practices.

– Survey of Public Resources, Models, and Tools: We summarize widely adopted open-source and commercial tools, as well as public models such as MusicLM [40] , Jukebox [41], MuseNet [42], Magenta [43], Riffusion [45], and AIVA [44]. The datasets, codebases, and platforms collected for this study are publicly available.

– Theoretical Understanding, Research Gaps, and Ethical Concerns: We synthesize theoretical foundations of AI-driven music intelligence, encompassing machine learning, deep learning, and computational creativity frameworks. Our analysis identifies key limitations in data diversity, cultural representation, interpretability, and creative authenticity. Furthermore, we discuss ethical challenges surrounding authorship, intellectual property, cultural appropriation, and the socioeconomic implications of AI in music production and education.

We first employ bibliometric methods to identify key research trends, influential works, and emerging clusters across the entire field, before conducting an in-depth systematic review that compares approaches, evaluates technological and pedagogical effectiveness, and identifies cross-domain synergies and challenges. This approach enables us to present both a macroscopic view of the field's development and a microscopic analysis of its most significant innovations, limitations, and challenges posed by AI's integration into musical practice [27, 50, 51].

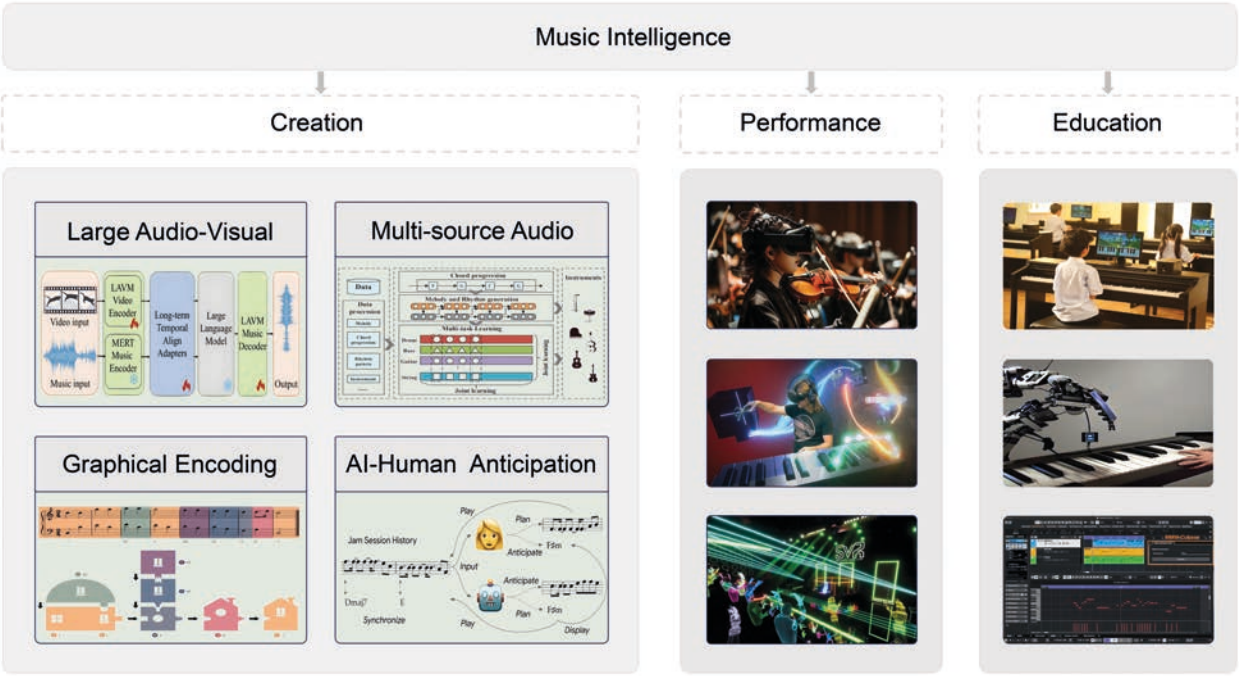

Figure 1. Key areas of music intelligence: creation, performance, and education

2 Methodology

In this section, we will present our review strategies and data collection methods.

2.1 Review Strategies

We first employ a comprehensive bibliometric analysis to map the evolving landscape of *music intelligence*. This analysis identifies core research trajectories, publication trends, and thematic shifts over time. We then conduct a systematic review of representative works that develop intelligent algorithms and technologies in three primary domains: music creation, music performance, and music education.

For the sake of review, our search strategy was carefully designed to ensure both breadth and precision. Primary search terms encompassed: 1) Music as the foundational disciplinary domain; 2) AI and its key methodological branches, including machine learning, deep learning, reinforcement learning, and neural networks; and 3) the three core application pillars – *Creation*, which includes melody composition and lyric generation; *Performance*, focusing on virtual reality (VR) technologies and digitalization that are widely used for improving performance; and *Education*, encompassing AI-powered pedagogical platforms, personalized education, computing and multimodal fusion. To ensure comprehensive retrieval while preserving thematic relevance, we systematically integrated domain-specific synonyms and related terminology into the query structure.

2.1.1 Bibliometrix Analysis

We employed Bibliometrix to assess research productivity and collaboration patterns. Key indicators (*e.g.*, annual publication growth rate, average citations per document, and international co-authorship rate) provided a macro-level overview of the research landscape. Additionally, Bibliometrix's timeline analysis can help trace the evolution of thematic focus areas over time, while journal analysis can reveal the most prolific sources contributing to AI-music research. These analytics can offer a comprehensive understanding of both the volume and distribution of scholarly output in this interdisciplinary field. To further explore the intellectual and collaborative structure of the field, we leverage VOSviewer [52, 46] to construct and visualize co-occurrence networks of keywords. The burst detection function in VOSviewer can identify keywords that exhibit sudden increases in citation frequency, highlighting emerging research topics and temporal trends.

2.1.2 Systematic Analysis

Key areas attracting substantial scholarly attention include AI-driven music creation, which investigates algorithms for autonomous and collaborative composition; music emotion recognition, which employs machine learning to analyze and classify affective content in audio signals; and applications in music performance and education, particularly in therapeutic contexts anchored by clinical and well-being outcomes. This systematic analysis is essential for synthesizing diverse findings across these domains, identifying research gaps, and mapping the evolution of key themes.

2.2 Data Collection

We sourced literature from three premier academic databases: Clarivate Web of Science (WoS), IEEE Xplore, and the ACM Digital Library, to ensure a comprehensive and high-quality dataset. These repositories were chosen for their extensive coverage of high-impact, peer-reviewed literature in technology domains, particularly due to IEEE and ACM's leading roles in publishing AI-related research relevant to music applications.

Figure 2 shows the process of study selection. A composite keyword combination was adopted: "AI music" or "music generation" or "music performance" or "music education" or "deep learning in music" or "artificial intelligence for music", resulting in a total of 1,112 initial records. The screening phase proceeded in two steps: first, through title and abstract screening, 302 records irrelevant to the "intersection of AI and music (core scenarios such as creation, performance, and education)" were excluded (*e.g.*, pure music theory research, art criticism without technical method support, or non-academic format content such as news and blogs); subsequently, full-text eligibility assessment was conducted on the remaining 810 literatures, further excluding three types of ineligible literatures, that is, 36 were excluded due to "incomplete technical/application evaluation indicators (*e.g.*, lack of quantitative data on model performance, empirical conclusions on application effects)", 21 were excluded due to "mismatch with target evaluation indicators (*e.g.*, focusing on traditional music tasks not driven by AI)", and 30 were excluded because "no deep learning or AI model was actually applied (while only basic audio processing techniques were involved)". During the screening process, should researchers disagree on the relevance or eligibility of literature, consensus shall be reached through arbitration by introducing new experts in the field, thereby ensuring the objectivity and consistency of screening outcomes. The finally included 723 literatures provide a core sample for subsequent bibliometric analysis and systematic review, comprehensively covering the technological evolution trajectory and diverse application scenarios in the field of Music Intelligence.

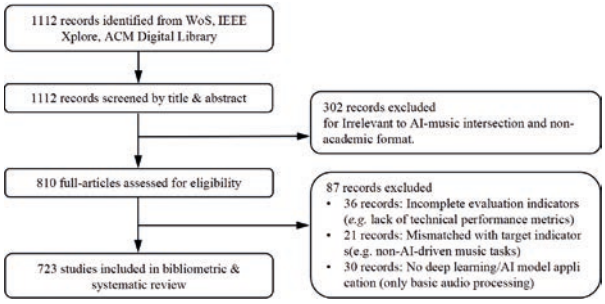

Figure 2. Process of study selection

As of May 12, 2025, we identified 485 music-intelligence-related publications from WoS, 404 from IEEE, and 223 from ACM. Among these, review papers accounted for 31 in WoS, 66 in IEEE, and 13 in ACM (see Figure 3). WoS includes 22 journal articles and 9 conference papers; IEEE includes 13 journal articles and 53 conference papers; and ACM includes 2 journal articles and 11 conference papers. All publications are indexed in databases such as Science Citation Index (SCI), SCI Expanded (SCIE), Social SCI (SSCI), Emerging Sources Citation Index (ESCI), and Arts & Humanities Citation Index (AHCI). The search followed a structured query framework tailored to the study's focus on AI applications in music. Only publications in English (*e.g.*, journal articles, review papers, and conference proceedings) were included.

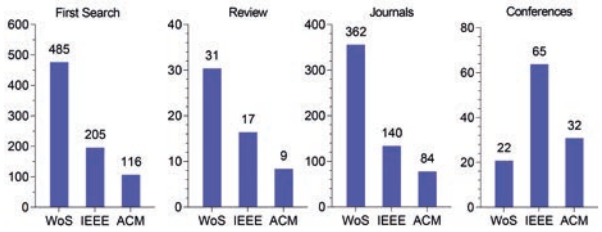

Figure 3. Distribution of music intelligence publications across major academic databases

3 Bibliometric Analysis

3.1 Statistics

Figure 4 summarizes the selected publications, which span 399 distinct journals and conference proceedings, highlighting the field's broad academic reach and inherently interdisciplinary nature. The dataset includes contributions from 8,649 authors, with only 81 single-authored works, emphasizing the dominance of collaborative research. On average, each publication involved 20.7 co-authors, and the international collaboration rate stands at 30.71%, reflecting strong global research ties. These publications have collectively 35,879 citations, averaging 14.88 citations per article, indicating notable academic impact. Analysis of 2,394 unique author keywords reveals a rich thematic landscape. The mean publication age of 4.36 years suggests that the field is both active and rapidly evolving.

The field of music intelligence demonstrated strong growth momentum, with an average annual growth rate of 14.92% (see Figure 5). Publication trends reveal two distinct phases: a gradual development period from 1991 to 2018, during which annual output remained below 20 papers, and a rapid acceleration phase beginning in 2019. This surge aligns with the rising demand for online education, particularly during the COVID-19 pandemic. Although a slight dip occurred in 2023, it is likely due to indexing delays rather than a real decline in research activity. Publication volume peaked in 2024 with 162 papers. The sustained growth affirms that AI-driven music research has evolved into a vibrant, interdisciplinary, and globally collaborative field.

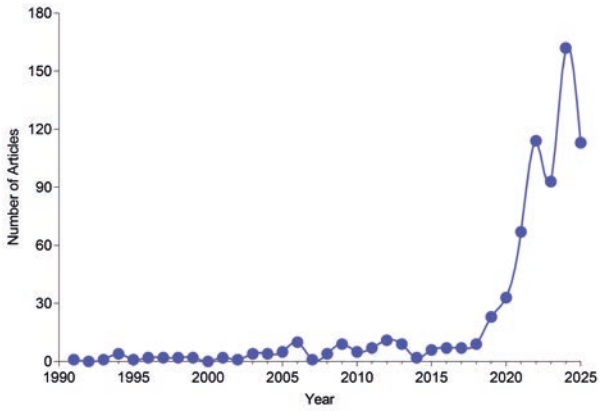

Figure 5. Trends in music intelligence research publications (1991-2025)

Figure 6 identifies the top ten journals by publication count in AI and music education research. Applied Sciences-Basel (23 articles) and IEEE Access (22 articles) lead as the most prolific outlets. Notable contributions also come from the EURASIP Journal on Audio, Speech, and Music Processing (17 articles), reflecting a signal processing focus, and Frontiers in Psychology (15 articles), emphasizing psychological perspectives. This diversity highlights the field's inherently multidisciplinary nature. Other prominent venues include the Journal of New Music Research, Wireless Communications and Mobile Computing, and AI Magazine, collectively illustrating the field's broad academic reach across computer science, audio engineering, psychology, and communication technologies.

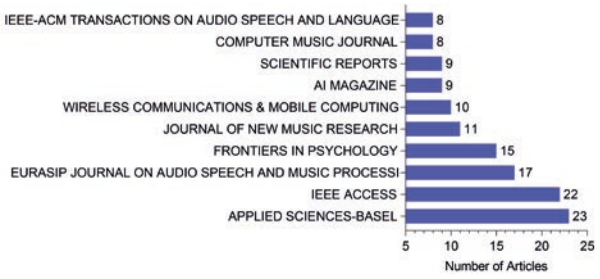

Figure 6. Statistical ranking of music intelligence publications by key academic journals

Figure 7 illustrates the international collaboration network. Among the 723 relevant publications (revised total), China leads with 188 publications (25.7%), reflecting strong research capacity and engagement. The United States follows with 118 publications (16.1%), and the United Kingdom ranks third with 81 publications (11.1%). Collectively, these three countries account for 52.9% of global output, underscoring their dominance in the music intelligence research landscape. China serves as a central hub, not only in publication volume but also in frequent international collaborations, particularly with the United States, the United Kingdom, Singapore, Germany, and India, forming a dense global network centered around China. Simultaneously, the United States and the United Kingdom act as key bridges within this network, fostering extensive academic partnerships across Europe and North America, notably with Canada, Italy, and Spain. This structure reflects a multilateral collaboration model anchored by China, the United States, and the United Kingdom, and highlights the field's inherently international and cooperative nature.

Figure 4. Bibliometric overview of music intelligence research in web of science

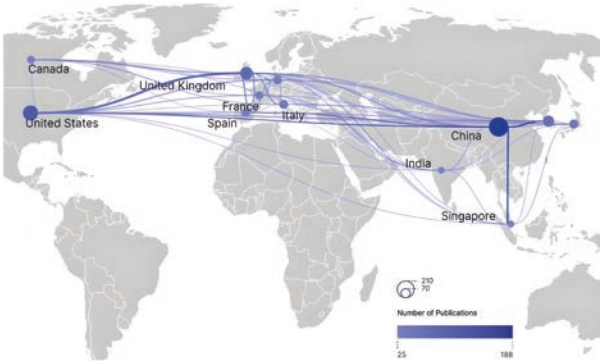

Figure 7. Global collaboration network in music intelligence research

3.2 Evolution and Trends

3.2.1 Keyword-based Research Topics

Analysis of author keyword evolution, shown in Figure 8, reveals a clear developmental trajectory within the music intelligence field, extending beyond the foundational prominence of artificial intelligence, deep learning, and machine learning. A significant trend is the technical evolution in modeling approaches. Early reliance on Long Short-Term Memory (LSTM) networks [53] and Hidden Markov Models (HMMs) has progressively given way to more advanced generative architectures, notably Transformer models and Diffusion Models. This shift signifies a move towards systems capable of more creative and autonomous musical content generation. Convolutional Neural Networks (CNNs) maintain a persistent role, underscoring their continued importance in core tasks like music feature extraction and rhythm analysis. The keyword landscape further highlights a pronounced diversification and specialization in application domains. Research increasingly targets specific verticals, particularly within healthcare and education. Growing frequencies of terms like anxiety, music therapy, and Electroencephalography (EEG) reflect an expanding focus on mental health applications, utilizing AI for emotion recognition and therapeutic intervention. Similarly, the association of music Education with Children indicates a rise in personalized AI-driven tools for learning and development. The emergence of VR and dance points towards exploration in immersive, multimodal experiences, where AI music generation integrates with movement and interactive environments. Interdisciplinary convergence is a defining characteristic of this evolution. Keywords demonstrate a deepening integration beyond core computer science and engineering. Sustained attention to creativity and emotion, alongside the appearance of Art and Synchronization, reveals efforts to bridge technological capabilities with artistic expression and human-centered design. This reflects a maturation of the field where research aims not only for algorithmic sophistication but also for meaningful engagement with human creative processes and affective responses. Music Information Retrieval remains a foundational element, enabling structured analysis of large-scale musical data for diverse applications, including commercial systems.

3.2.2 Clustering-based Analytics

We conducted a clustering analysis of keyword co-occurrence patterns, visualized as a keyword knowledge graph in Figure 9. Sixteen distinct clusters emerged, representing thematic groupings such as 1) AI, music generation, and speech recognition; 2) music, algorithms, and sensors; 3) performance, generative models, and adaptation; and 4) machine learning, algorithmic composition, and ap-

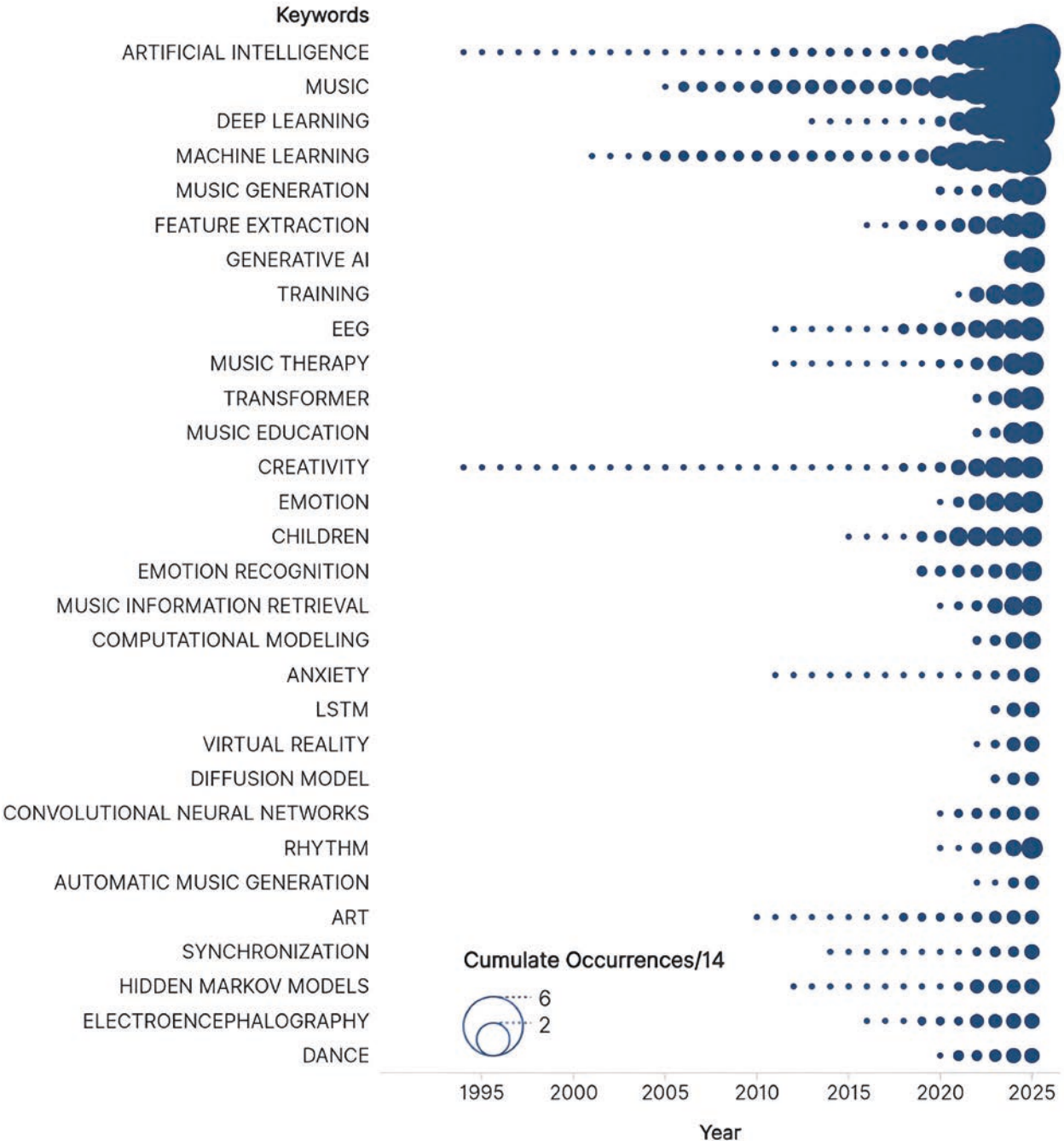

Figure 8. Dominant keywords in music intelligence research over three decades

plications. Keywords like target recognition, sensors, and data point to an expansion into computer vision, IoT systems, and multimedia technologies. This trend suggests a growing research focus on multimodal perception systems, where AI music interacts dynamically with environmental and visual inputs. Emerging applications include smart spaces and responsive installations, with IoT integration paving the way for ambient musical intelligence driven by real-time sensor data. Looking ahead, AI music systems are expected to integrate seamlessly with smart environments, anticipating user moods and activities to generate adaptive, context-aware soundscapes that transform everyday spaces into responsive sonic ecosystems.

3.2.3 Emerging Trends and Future Directions

Deep learning and machine learning remain the foundations of AI music research, connecting diverse subfields from generation to emotion recognition and therapeutic applications. Advances in neural networks and co-attention mechanisms have enabled more precise modeling of musical affect and context, bridging algorithmic design with real-world use. Future models are expected to achieve heightened emotional and contextual awareness, powering hyper-personalized systems that adapt to users' physiological and behavioral cues. A clear shift toward socially impactful applications is emerging. The prominence of keywords such as anxiety, intervention, and emotion signals growing interest in therapeutic uses of AI-generated music for emotional regulation and psychological support. By integrating emotion perception with generative algorithms, researchers are developing adaptive soundscapes for personalized therapy and preventive mental health care. Emerging trends that are highlighted by VR, synchronization, and automatic music generation point to immersive, interactive, and context-responsive systems. These innovations promise to transform entertainment, experiential art, and therapeutic design through adaptive sonic environments that evolve with human presence and emotion.

4 Systematic Review

The following sections provide a systematic review of AI technologies applied across music creation, performance, and education.

4.1 Creation Intelligence

AI is fundamentally reshaping music creation through two interrelated domains: melody composition and lyric generation. As outlined in Table 1, recent algorithmic advancements have driven these domains beyond basic automation toward sophisticated functionalities such as stylistic control, human-AI co-creation, and multimodal interaction.

4.1.1 Melody Composition

The evolution of AI-driven melody composition since the early 21st century has unfolded across three distinct phases: rule-based systems, deep learning models, and contemporary multimodal approaches [8]. Early research primarily focused on tasks such as automatic accompaniment and melody completion. Current efforts, however, emphasize advanced capabilities including structural modeling, stylistic control, polyphonic coordination, and enhanced human-AI interaction. This trajectory reflects a dual progression toward increasing technical sophistication and creative democratization.

Notable advancements illustrate this evolution. For instance, notation-integrated composition frameworks like the Measure-by-Measure system [54] enable bar-level automatic composition using modern staff notation. Its innovative grid-based design precisely models temporal dependencies and part alignment, facilitating seamless integration of traditional musical notation with AI algorithms while maintaining high compositional accuracy.

The interactive dimension has been further enriched through immersive systems leveraging virtual reality and spatial interaction technologies. Such platforms allow composers to trigger sound events in real time via handheld controllers, dynamically manipulating musical elements within a 3D environment [55]. Cross-modal generative models mark another cutting-edge development. The Video-to-Music approach [11] employs large language models (LLMs) enhanced with advanced memory mechanisms for video-driven music generation. This method overcomes limitations in processing short video segments and enables the efficient creation of long-sequence music tightly synchronized with visual content.

Additionally, controllable hierarchical architectures like MIDI-GPT [12], a Transformer-based

Table 1. AI technologies for melody composition and lyric generation

	Authors & Year	Technology	Keywords
Melody	Yan and Duan (2024)	Bar-level automatic composition framework based on staff notation	• Grid Structure • Temporal Dependency • AI Algorithm Fusion
	Tomasetti and Turchet (2024)	Immersive music creation driven by virtual reality	• Virtual Reality • Sound Event Triggering • Human-AI interaction
	Kai and Xing (2024)	Long-sequence music generation driven by video	• Generative Model • Memory Enhancement • Video Understanding
	Zhang et al. (2024b)	Application of AI tools in music education	• Composition Ability • Educational Reform • Teaching Empowerment
	Pasquier et al. (2025)	Controllable generation model based on Transformer	• Conditional Filling • Instrument Control • Structural Consistency
	Pu et al. (2025)	Music creation and education platform for non-professional users	• Generative Creation • Hierarchical Creation • Interest in Learning Music
Lyric	Navarro-Caceres et al. (2020)	Melody-driven lyric generation system	• Markov Model • Rhythm Consistency • Emotion Matching
	Ma et al. (2021)	Lyric generation pipeline based on melody structure analysis	• MIDI Input • Syllable Distribution • Stylistic Features
	Vechtomova et al. (2021)	Real-time lyric generation for live performances	• Adversarial Spatial • Topological Mapping • Audio-text Collaboration
	Ma et al. (2024)	Keyword-driven hierarchical coherent lyric generation	• Graph Structure • Coherence Mechanism • Vocal Learning

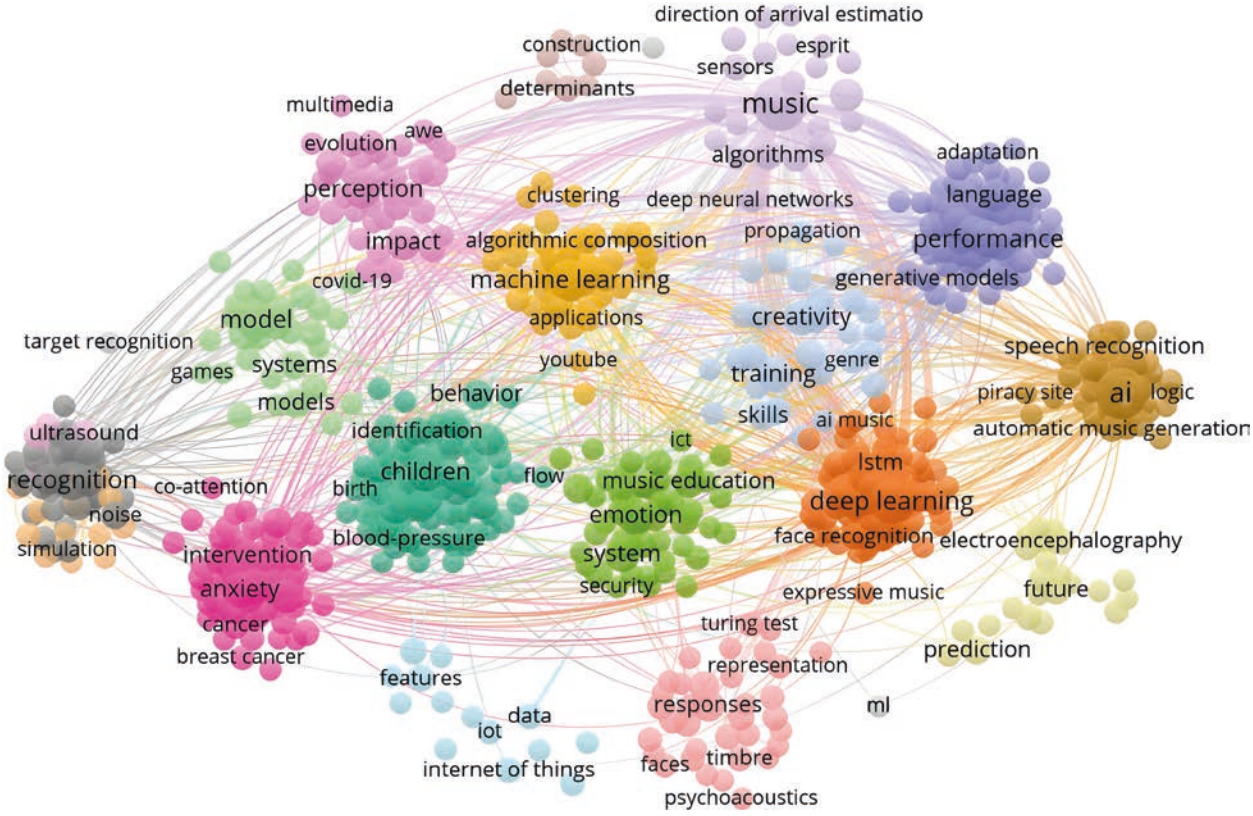

Figure 9. Keyword co-occurrence network and clustering patterns in music intelligence research

model, offer conditional completion at both track and bar levels. Supporting multi-attribute control over instruments, style, and polyphony, MIDI-GPT enhances the structural coherence of generated compositions while empowering users with flexible creative constraints.

4.1.2 Lyric Generation

Lyric generation technology has evolved substantially along three main trajectories: transitioning from melody-driven to semantics-driven approaches, advancing from static text generation to dynamic human-computer interaction, and progressing from simple mapping techniques to deep semantic modeling. These developments have significantly enhanced lyrical expressiveness in logic, rhythm, and emotional conveyance. Generative language models and multimodal systems now lead the research frontier [41], enabling sophisticated semantic alignment that surpasses basic music-language synchronization.

Key innovations illustrate this progression. Early systems like ETHNO-MUSIC [58] employed Markov models to generate Spanish popular melodies and integrated the Tra-La-Lyrics module to maintain rhythm and emotional consistency between melody and lyrics, achieving positive subjective evaluations through fundamental alignment. Subsequent efforts introduced structured pipeline architectures. AI-Lyricist [59] features a four-module pipeline combining MIDI inputs with user-provided keywords. It sequentially performs melody structure analysis, lyric generation via SeqGAN with multi-round adversarial training, prosody alignment, and semantic refinement, producing lyrics noted for logical coherence, structural clarity, and stylistic consistency.

Real-time interactive platforms have emerged to support live lyric creation. [60] targets improvisational performance by aligning latent spaces of audio and text through adversarial space alignment and topological mapping, offering musicians immediate lyrical inspiration and arrangement suggestions during live sessions.

Recent advances focus on deep semantic understanding. [61] achieved breakthroughs through unsupervised keyword skeleton extraction, graph-based multi-layer semantic expansion, and coher-

ence mechanisms across lexical, syntactic, and discourse levels. This approach improved objective coherence scores by 5% and subjective ratings by 19%, demonstrating applications such as generating language learning content for singing education.

4.1.3 Network Architectures

Although an in-depth technical review of neural architectures lies beyond the scope of this work, we present a synthesized overview of the principal architectural paradigms applied in automatic music generation, as illustrated in Table 2 and Table 3. These include Generative Adversarial Networks (GANs) [36], Variational Autoencoders (VAEs) [62], Transformer models [34], evolutionary algorithms [63], and rule-based systems [64]. Each category brings distinct strengths to the task of composing music, from enhancing realism to ensuring stylistic consistency or interpretability. The following describes the architectures.

- GANs have emerged as a prominent approach due to their capacity to generate highly realistic musical outputs that are often indistinguishable from those in the training data. Comprising a generator and a discriminator engaged in an adversarial learning process, GANs gradually improve their generative capabilities through continuous feedback. Recent GAN-based systems [91, 95] have incorporated elements such as emotional conditioning, latent space manipulation, and diffusion-based decoding. Hybrid approaches that integrate VAEs with GANs (such as those proposed by [93] and [92]) demonstrate that combining architectural paradigms can enhance both controllability and expressiveness in musical output. Out of the 118 systems surveyed, 19 employed GANs, often in combination with recurrent neural networks or feedforward architectures, to boost sequence learning capabilities.

- VAEs, on the other hand, provide a structured latent space in which musical representations can be encoded and decoded, enabling interpolation and transformation of musical features [96]. Their ability to capture high-level variations in data makes them particularly suitable for creative manipulation. A well-known example is MusicVAE [99], which uses a hierarchical structure to support long-term temporal coherence. Other representative works include [35] on interpolative composition and [41] on raw audio compression. More recent contributions, such as the multi-source latent diffusion strategy by [10], highlight how VAEs can be integrated with other generative paradigms to further expand their creative potential. Among the systems examined, 18 utilized VAEs either as standalone architectures or in conjunction with other models for enhanced stylistic control.

- Transformer-based architectures have shown considerable promise in sequence modeling due to their attention mechanisms and ability to handle long-range dependencies. These models are particularly effective in capturing hierarchical structures and complex rhythmic relationships in music. Notable examples include systems that focus on multi-instrumental generation [68], chord-conditioned melody generation [23], and emotion-driven outputs [22]. Furthermore, some Transformer-based architectures are combined with GANs [65], blending structural modeling with adversarial refinement.

- Inspired by principles of natural selection, evolutionary algorithms treat musical ideas as evolving entities. In these systems, musical fragments act as chromosomes, with genes representing notes, chords, or rhythmic motifs. Mutation and crossover operators are applied to generate variations, while a fitness function evaluates their aesthetic quality. Over successive generations, this leads to musical outcomes better aligned with the target style. For instance, [84] use pitch-based gene encoding, and [88] apply genetic evolution techniques to traditional Chinese music. Although only 13 systems in the corpus are purely evolutionary in nature, this approach is frequently embedded in hybrid models, *e.g.*, [85], often in combination with CNNs and RNNs to balance creative diversity and learned structure.

- Rule-based systems offer a different paradigm, relying on explicit symbolic rules or music-theoretical constraints to guide the generation process. While these systems may lack the flexibility of deep learning-based models, they offer greater interpretability and stylistic consistency. For example, [104] introduced a system based on reward-guided rule constraints, while [105]

Table 2. Network architectures (transformer-based and evolutionary approaches) for music creation

	GAN	Authors (Year)	Technologies
Transformer	YES	Muhamed et al. (2021)	Adversarial Long-Sequence Music Generation
		Jin et al. (2022)	Transformer with Music-Theory Reward Optimization
		Neves et al. (2022)	Emotion-Driven Transformer-GAN
		Wang et al. (2024)	Style-Conditioned Transformer Generation
		Lan et al. (2024)	Text-to-Music and Temporal Conditioning
		Pasquier et al. (2025)	Attribute-Controlled Transformer Music Generation
	NO	Huang et al. (2018)	Linear Relative Attention Music Generation
		Donahue et al. (2019)	Multi-Instrument Transformer Transfer Generation
		Payne (2019)	Multi-Style Multi-Instrument Transformer Generation
		Ens and Pasquier (2020)	Multi-Track Segmented Sequence Generation Control
		Huang and Yang (2020)	Hierarchical Rhythmic Structure Data Representation
		Hadjeres and Crestel (2021)	Structured Encoding Linear Transformer Inpainting
		Choi et al. (2021)	Chord-Conditioned Two-Stage Melody Generation
		Makris et al. (2021)	Valence-Conditioned Lead Sheet Generation
		Zhou et al. (2023)	Relative Positional Attention
		Luo et al. (2024)	Spatiotemporal Features and Multitrack Modeling
		Gao et al. (2024)	Theme Variation Extraction and Generation
Evolutionary	YES	Zhu et al. (2022)	Multimodal Conditional VQ Audio Generation
		Han et al. (2024b)	Guzheng Motif Modeling and Residual LSTM
	NO	Jeong et al. (2017)	Multi-objective Evolutionary Melody Generation
		Lopes et al. (2017)	Multi-objective Evolutionary Creative Composition
		De Felice et al. (2017)	Automatic Four-Voice Composition
		Masuda and Iba (2018)	RNN and Interactive Evolutionary Computation
		Mo et al. (2018)	Simulated Annealing Genetic Algorithm Composition
		Stoltz and Aravind (2019)	Music-Psychology Genetic Algorithm Composition
		Wen and Ting (2020)	Bossa Nova Genetic Algorithm Composition
		Shi and Wang (2020)	Genetic Neural Network Assisted Composition
		De Prisco et al. (2020)	Multi-objective Evolutionary Four-part Harmony
		Sabitha et al. (2021)	Multi-Algorithm Fusion Music Generation
		Zeng and Zhou (2021)	Memetic Algorithm for Traditional Music Generation
		De Azevedo Santos et al. (2021)	Genetic Algorithm Emotion-guided Composition
		Kilb and Ellis (2024)	Evolutionary Algorithm Human-AI Co-Creation

Table 3. Network architectures (VAE and rule-based systems) for music generation

	GAN	Authors (Year)	Technologies
VAE	YES	Qiu et al. (2019)	Emotion-driven Music Generation
		Cheng et al. (2020)	Latent Space Music Generation
		Wang et al. (2020)	VAE-GAN Music Generation
		Huang and Huang (2020)	Emotion-Driven Automatic Composition
		Cheng et al. (2020)	Innovative Latent Structure in Temporal GANs
		Zhang et al. (2023)	Emotion-conditioned Diffusion VAE-GAN Generation
		Lam et al. (2023)	Semantic-Guided Diffusion Decoding
		Gan (2024)	VAE + Diffusion Emotion-Adaptive Generation
	NO	Tikhonov et al. (2017)	Temporal Complex and Diverse Generation
		Brunner et al. (2018)	Multi-Track Music Style Transfer
		Roberts et al. (2018)	Hierarchical Decoding to Prevent Posterior Collapse
		Masuda and Iba (2018)	Integration of RNN and Interactive Evolution
		Lattner and Grachten (2019)	Unsupervised Coding for Kick Drum Generation
		Hung et al. (2019)	Jazz Transfer Learning Generation
		Jia et al. (2019)	Coupled Latent Variables with Binary Regularization
		Dhariwal et al. (2020)	Multiscale VQ-VAE Audio Generation
		Grachten et al. (2020)	Temporally Stable Two-Dimensional Latent Space
		Lim et al. (2020a,b)	Continuous Style Embedding Variational Autoencoder
		Diéguez and Soo (2020)	VAE-Based Music Interpolation Generation
		Grekow and Dimitrova (2021)	Emotion-Conditioned Variational Autoencoder
		Zhang et al. (2024a)	Discrete Diffusion Symbolic Music
		Xu et al. (2025)	Multi-source Latent Diffusion Generation
Rule-based	YES	Jin et al. (2020)	Reward-feedback Constrained Composition
		Zhu et al. (2024)	Discrimination and Novel Symbolic Representation
	NO	Manzelli et al. (2018)	Symbolic and Raw Audio Joint Generation
		Wiriyachaiporn et al. (2018)	LSTM and Rule-based Algorithms
		Cunha et al. (2018)	Integer Programming Optimized Guitar Improvisation
		Huang et al. (2024)	Non-differentiable Rule Diffusion Guidance
		Tian et al. (2025)	Multimodal Symbolic Music Generation

integrated symbolic representation with neural networks. In some cases, rule-based layers are embedded within LSTM or Transformer backbones, *e.g.*, [107], enabling a fusion of symbolic control and learned generalization.

A cross-analysis reveals that recurrent neural networks-particularly LSTM variants – still form a foundational component of many systems, although not explicitly featured. Overall, GANs and VAEs are valued for their flexibility and generative richness, Transformers are gaining ground due to their structural modeling prowess, evolutionary algorithms provide an intuitive framework for stylistic exploration, and rule-based approaches continue to play a vital role in ensuring harmonic and theoretical integrity. Many state-of-the-art models leverage multiple architectural paradigms (*e.g.*, Transformer-GANs) to combine their respective strengths, achieving both expressive depth and stylistic fidelity in music generation.

4.1.4 Public Resources

The rapid advancement of AI in music creation has been significantly propelled by the open-source community and the emergence of specialized platforms, providing essential models, tools, and infrastructure, such as Soundraw [110], Amper Music [111], AIVA [44], Ecrett Music [112], and Boomy [113], as Table 4 shows.

These platforms address diverse needs, including melody generation, style control, lyric composition, and commercial music production, significantly expanding the practical application boundaries of AI in music creation. For example, as shown in Table 5, MuseGAN [115] introduced the first multi-track music generation framework, supporting polyphonic creation from scratch or automatic accompaniment for existing tracks, pioneering multi-part collaboration using GANs. MelGAN [116] offered a non-autoregressive feedforward convolutional architecture, successfully training GANs to generate audio waveforms without distillation or perceptual loss, providing efficient reconstruction for text-to-speech and unconditional music generation. A significant step towards interpretability came with DDSP (Differentiable Digital Signal Processing) [117], integrating traditional DSP methods with neural networks for controllable features like pitch, loudness, and timbre, enabling high-fidelity, modular systems. [41] introduced a paradigm shift by modeling raw audio directly using multi-scale VQ-VAE encoding and autoregressive Transformer decoders, capable of generating high-fidelity, long-form music with vocals, controllable by lyrics, singers, and styles. Innovation continued with [45], applying diffusion models to music generation via spectrograms, enabling real-time creation suitable for interactive and live performance contexts. By 2024, models demonstrated enhanced expressiveness and detailed instrument modeling. ViolinDiff [15] focused specifically on violin audio synthesis, employing a two-stage diffusion model to capture natural fundamental frequency contours and pitch bends. Esteso [17] explored AI-impro interaction, specializing in recognizing and reproducing extended bass playing techniques to enhance expressiveness. In 2025, the open-source foundational model YuE [118] was released, designed for lyric-to-song generation, producing complete musical works minutes in length including lead vocals and accompaniment across multiple styles and vocal techniques, noted for high quality and generalization.

4.2 AI in Music Performance

With the integration of AI technologies such as Virtual Reality (VR) and Augmented Reality (AR), musical performance characterized by immersion, interactivity, and digitalization is shown in Table 1.

4.2.1 VR and Interactive Performance

The convergence of VR for musical performance has catalyzed significant innovation in establishing direct mappings between bodily movement and sound, profoundly enhancing the potential for immersion and interactivity on stage. Initial feasibility studies, exemplified by Yamaha's experimental system [119], demonstrated the real-time control of piano performance through body gestures, thereby providing crucial validation for the core concept of gesture-to-acoustic mapping. Building directly upon this foundation of gesture control, the "Human-AI Duet" system [16] integrated advanced music tracking, pose recognition, and AI-generated virtual performance motions. This integration enabled a virtual violinist to autonomously accompany a human pianist, creating an expandable framework for multimodal interactive performance. Subsequent efforts

Table 4. Open-Platform AI tools for music creation

Authors (Year)	Tool (Link)	Keywords
Ecrett Music (2018)	Ecrett Music (https://ecrettmusic.com/)	• Auto Background Music Generation • Soundtrack for Video • Mood-based Generation
Boomy Corporation (2023)	Boomy (https://www.boomy.com/)	• Fast Music Generation • Social Sharing • Style Selection
Soundraw (2025)	Soundraw (https://soundraw.io/)	• Automatic Melody Generation • Royalty-free Music • Style Customization
ShutterstockInc. (2025)	Amper Music (https://www.shutterstock.com/discover/amper-music)	• Commercial Music Generation • Style Control • AI Composition
Aiva Technologies SARL (2025)	AIVA (https://www.aiva.ai/)	• AI Composition • Film Scoring • Style Imitation
AI Tool Selection (2025)	Music Toolkit (https://aitoolselection.com/zh-CN)	• Aggregated Generation Platforms • Copyright Music Index

focused on enriching the performance environment, culminating in the "VRhythm" project [120]. This initiative developed dedicated virtual stages, crafting multisensory and dynamically adjustable audiovisual environments specifically designed to amplify performers' expressive capabilities and deepen emotional communication within the musical experience. A pivotal shift towards active generation occurred with the MoMusic system [9].

4.2.2 Digitalization

Digitalization in music is being propelled by augmented reality (AR) and AI perception technologies, enabling both the revitalization of traditional instruments and the development of inclusive performance systems. A prominent example involves the digital reconstruction of China's ancient bianzhong (chime bells), where an interactive system integrating motion sensing and AR projection transformed static museum artifacts into dynamic, playable installations. This allowed audiences to trigger authentic bell tones through body movements while receiving real-time visual feedback conveying cultural semantics, creating an integrated "sound-motion-cultural meaning" experience. Empirical results confirmed significantly enhanced participant immersion and cultural engagement, demonstrating how such technologies facilitate "musical regeneration" of cultural heritage while converging art, education, and technology [121]. Meanwhile, extending beyond heritage revitalization, digitalization drives accessibility innovation through systems like the networked Accessible Digital Musical Instruments (ADMIs) studied by [122]. Designed for collaborative performances involving teaching assistants, students with Profound and Multiple Learning Disabilities (PMLD), and professional musicians, mixed-methods evaluation revealed that these instruments reliably supported musical contexts without technical disruption.

4.3 AI for Music Education

AI serves as a catalytic force driving structural transformations in music education, fundamentally reshaping pedagogical paradigms. A comprehensive bibliometric analysis spanning 1991-2024 establishes AI's pivotal role in enhancing instructional interactivity, enabling personalized learning pathways, and promoting educational inclusivity [123]. This large-scale study identifies key technological trajectories, research hotspots, and geographical distributions, providing empirical foundations for under-

Table 5. Open-Source AI models for music creation

Authors [Year]	**Model (Open Source)**	**Keywords**
Dong et al. (2018)	MuseGAN (https://salu133445.github.io/musegan)	• Multi-track Music Generation • Multi-part Coordination • GAN, Polyphonic Generation
Kumar et al. (2019)	MelGAN (https://github.com/descriptinc/melgan-neurips)	• Audio Waveform Generation • Non-autoregressive • Feed-forward Convolution • Text-to-speech
Engel et al. (2020)	DDSP (https://github.com/magenta/ddsp)	• Digital Signal Processing • Timbre Transfer • Pitch Control
Dhariwal et al. (2020)	Jukebox (https://jukebox.openai.com/)	• Vocal Music Generation • VQ-VAE, Transformer • Lyric Conditioning • Large-scale Music Modeling
Forsgren and Martiros (2022)	Riffusion (https://www.riffusion.com/)	• Real-time Music Generation • Spectrogram-based • Diffusion Model • Interactive Creation
Kim et al. (2024)	ViolinDiff (https://github.com/daewoung/ViolinDiff)	• Violin Audio Modeling • Fundamental Frequency Contour • Pitch Bending
Stefani et al. (2024)	Esteso (https://github.com/domenicostefani/Esteso)	• Interactive Improvisation • Double Bass Generation • Extended Technique Modeling
Yuan et al. (2025)	YuE (https://github.com/multimodal-art-projection/YuE)	• Lyric-to-Song Generation • Vocal and Accompaniment Synthesis • Chinese Traditional Style

Table 6. AI-enhanced music performance technologies: virtual reality and traditional instrument

	Authors (Year)	Technologies	Keywords
Virtual Reality	Zulić (2019)	Motion-controlled piano performance	• Body-sound Mapping • Gesture Recognition • Interactive Control
	Lin et al. (2020)	Automatic co-performance of virtual violin and human piano	• Pose Recognition • Music Tracking • Virtual Performance
	Ppali et al. (2022)	Multisensory immersive VR performance environment	• Visual-auditory Space • Emotional Expression • Motivation-driven • Multimodal Experience
	Bian et al. (2023)	Real-time motion-driven music generation	• Hand Tracking • Rhythm Mapping • Virtual Timbre Control • Bodily Composition
Traditional Instruments	Guo et al. (2023)	Traditional instrument performance system based on AR and motion sensing	• Bianzhong Performance • Motion Recognition • Cultural Projection • Augmented Reality • Museum Interaction
	Lindetorp et al. (2023)	AR-Somatosensory Interactive System for Traditional Instrument Performance	• Music-making • Web Audio • Accessible Instruments • Interactive Systems

standing how AI reconfigures educational philosophies and pedagogical approaches across global contexts. Table 7 presents the typical AI platforms for music education, personalized education, inclusive computing, assessment, etc.

4.3.1 AI-Powered Platforms

The evolution of AI-powered music education platforms has progressed from rudimentary systems to sophisticated learner-centered ecosystems, characterized by sequential technological breakthroughs. The foundational stage emerged in 2018 with virtual collaborative composition environments, where studies observed autonomous student learning behaviors through iterative experimentation cycles. These platforms pioneered behavioral analytics for content adaptation, thereby establishing initial frameworks for self-regulated learning despite technological limitations [124]. Building upon this groundwork, significant architectural advancement followed in 2022 with an AI vocal training system for music majors, integrating expert systems, neural networks, and speech emotion recognition. Its four synergistic modules-course management, real-time error correction, self-assessment, and adaptive intervention collectively demonstrated measurable gains in engagement through emotion-sensitive pedagogy [125]. A transformative leap then occurred in 2024 with the SONATA platform, which implemented Transformer-based multi-task learning to simultaneously generate affect-driven musical content and align teaching strategies with skill levels. Crucially, its integrated emotion recognition module enabled granular analysis of musical expression, creating biofeedback loops for aesthetic development [4]. Most recently, the frontier has extended to Bi-LSTM networks with attention mechanisms for MIDI performance analysis, achieving an interpretable, granular assessment in piano pedagogy. While demonstrating operational robustness, this approach nonetheless requires expansion to diverse instruments for broader applicability [33].

4.3.2 AI-Enabled Personalization

The evolution of AI-enabled personalization in music education has progressively shifted from standardized instruction toward precision pedagogy, marked by increasingly adaptive, data-driven, and culturally responsive systems. This trajectory originated with the SONATA intelligent tutoring system [126], which established foundational principles through multi-strategy adaptation, integrating exploratory learning, dynamic scaffolding, and cognitive diagnosis to deliver individualized instruction based on student profiles. Building upon these early frameworks, subsequent decades leveraged generative AI, chord recognition, and transcription technologies to create student-centered platforms. Notable examples include RealEarTrainer's interest-aligned auditory routines and AI-assisted piano systems that auto-generate technique exercises matching repertoire difficulty, thereby enhancing pedagogical relevance while accommodating stylistic diversity [18]. Concurrently, commercial platforms [31] like SmartMusic and Yousician advanced educational equity through algorithmic learning trajectories and accessibility features. The current frontier features – Deep Fuzzy Music Tutor framework (DFMT), synthesizing deep learning with fuzzy logic for nuanced real-time evaluation. By extracting multidimensional performance features, it delivers feedback aligned with pedagogical traditions like Kodály and Orff methods [5]. Furthermore, contemporary systems increasingly embed style imitation and genre modeling to enable culturally responsive instruction, though researchers emphasize these tools augment rather than replace human artistic judgment [133].

4.3.3 Multimodal Fusion

Affective computing and multimodal fusion have profoundly reshaped music education by introducing emotionally responsive and immersive learning experiences. The foundations of affect-aware pedagogy emerged in 2020, when computational models began mapping lyrics to emotions using affective lexicons such as the Affective Norms for English Words (ANEW), thereby establishing semantic–emotional scaffolds for educational applications [127]. Building on this groundwork, subsequent frameworks linked cognitive and emotional development, identifying synergies between music education and sports to propose emotion engagement-centered learning models [128]. Social robotics further extended affective pedagogy, where sensor-equipped robots modulated their responses through interactive electronic music, offering adaptive emotional feedback [25].

Table 7. AI-driven technological innovations in music education: applications across pedagogical domains

	Authors (Year)	Technologies	Keywords
CITP	Biasutti (2018)	Behavior Feedback	• Self-directed Learning • Collaborative Composition • Constructive Learning
	Wang (2022)	Expert System	• Personalized Recommendation • Alleviating Teacher Shortage • Dynamic Path Adaptation
	Chen and Sun (2024)	Transformer	• Emotion-driven Generation • Strategy Auto-adjustment • Emotional Embedding
PIT	Angelides and Tong (1995)	Exploratory Learning	• Tailored Instruction • Cognitive Identification
	Sanganeria and Gala (2024)	Generative AI	• Auditory Training • Learning Engagement
	Chen (2025)	Deep Learning	• Multidimensional Extraction • Kodály Method Integration
EU HAIC	Ara and Gopalakrishna (2020)	Text Analysis	• Emotion Mapping • Semantic Sentiment
	Shi (2023)	Deep Learning	• Critical Thinking • Emotional Regulation • Teamwork Collaboration
III	Sun (2024)	VR Classroom	• Visual Interaction Analysis • Learning Outcome Prediction
	Han et al. (2024a)	Mapping Features	• Synesthetic Perception
	Ying (2025)	Voice Robot	• Personalized Classroom • Enhanced Engagement • Enhanced Visual Aesthetics
EI for SG	Vargas et al. (2020)	VR Rehabilitation	• Cognitive Impairment • Higher Participation
	Pérez et al. (2024)	VR Visualization	• Social Presence • Memory Stimulation

Parallel to these pedagogical advances, multimodal integration emerged as a critical driver of engagement and personalization. VR-based learning environments enhanced spatial-temporal skill acquisition [129], while intelligent teaching terminals combined speech recognition, immersive interfaces, and deep learning-based emotion adaptation to create responsive online classrooms [19]. Adaptive systems leveraging computer vision for real-time behavioral recognition further refined this interactivity, dynamically adjusting pedagogy and employing KNN-based music recommendation for personalized content delivery [134, 135]. The MusicColors system extended these ideas through synesthetic interaction loops of "emotion perception-visual feedback-creative stimulation", providing therapeutic and creative benefits for diverse learners [136].

From a computational standpoint, Affective Music Generation (AMG) systems have evolved from rule-based and data-driven approaches to hybrid paradigms, yet continue to face challenges in emotion labeling and evaluation [137]. Advances in multimodal emotion recognition have addressed these limitations by integrating lyric-audio fusion, outperforming acoustic-only methods [138]. Similarly, the bimodal deep model proposed by [139] demonstrated that combining multiple information sources more accurately captures the emotional dimensions of music.

As demand for multi-dimensional artistic expression in Music Intelligence (MI) grows, multimodal architectures have become central to both technological and educational innovation. Music inherently integrates acoustics, vision, motion, and semantics. Dimensions often fragmented in single-modal systems such as LSTM-based melody generators relying solely on audio. Multimodal frameworks overcome these constraints through three mechanisms: (1) cross-modal feature encoding, transforming data from diverse modalities (*e.g.*, lyrics, gestures, or images) into a unified representational space; (2) inter-modal semantic alignment, mapping meanings across modalities (*e.g.*, linking "sad visuals" with "minor harmonies"); and (3) collaborative generation, enabling mutual reinforcement among modalities for greater expressiveness. Representative systems include MoMusic [9], which fuses motion and audio using a Transformer-based cross-modal encoder to align rhythm, amplitude, and spatial motion features, and the video-driven music generation model of [11], which employs cross-modal attention to synchronize scene semantics and acoustic features for emotionally coherent compositions. Despite their promise, these architectures face ongoing challenges, including ambiguous cross-modal semantic mapping (stemming from the lack of unified cognitive standards) and increased computational complexity, which creates trade-offs between expressiveness and real-time performance.

4.3.4 Inclusive and Therapeutic Applications

The scope of AI music education platforms has expanded significantly beyond pedagogical enhancement to encompass inclusivity and therapeutic interventions, demonstrating their capacity to address diverse social, cultural, and cognitive needs, particularly among marginalized populations. This broadening trajectory began with foundational research in 2020 that pioneered VR systems for cognitive rehabilitation. Using an action research framework, researchers co-developed a configurable musical VR application that improved memory, coordination, and spatial perception while simultaneously boosting emotional responsiveness and engagement, thereby establishing a vital model for therapeutic technology adaptation [131]. Subsequently, academic focus extended to sociocultural accessibility dimensions, exemplified by a landmark 2023 analysis of eleven Accessible Digital Musical Instruments (ADMIs) for deaf/mute communities. Grounded in medical, social, and cultural disability models, this work emphasizes that inclusive design must transcend physical accommodations to address social participation and cultural identity needs, providing essential theoretical foundations for socially responsive music technologies [140]. Parallel developments in 2024 advanced therapeutic applications through the Comodia Elderly Project, where 180-degree stereoscopic VR performances enhanced spatial presence, emotional arousal, and autobiographical memory recall in elderly individuals with intellectual disabilities, highlighting the therapeutic potential of immersive technology in geriatric care [132]. Simultaneously, researchers addressed auditory exclusion through synesthetic typographic modulation, translating musical attributes like loudness and rhythm into visual typography changes. This innovation created alternative music comprehension

channels for the hearing-impaired by evoking cross-modal sensory experiences [130].

4.3.5 Intelligent Assessment and Feedback

The integration of artificial intelligence in music education has evolved toward intelligent assessment and adaptive feedback mechanisms, transforming traditional evaluation into dynamic systems that track cognitive and expressive development. This progression began with foundational 2015 research on online collaborative composition environments, where behavioral analysis revealed that pedagogically structured tasks and platform design could effectively stimulate metacognitive awareness, providing critical insights for AI-assisted assessment systems [141]. Building upon these principles, 2022 witnessed systematic machine learning adoption in performance evaluation. Comparative studies demonstrated that Gradient Boosting Decision Trees (GBDT) outperformed logistic regression, SVM, and random forests in accuracy and robustness for modeling student data, confirming AI's potential to enhance feedback reliability [26]. Concurrently, the Fuzzy Analytic Hierarchy Process (Fuzzy AHP) quantitatively evaluated five AI paradigms in music gaming environments, establishing frameworks for assessing interactivity and instructional efficacy [142]. Subsequent developments refined technical approaches: a 2023 conceptual study proposed integrating teaching simulations with data analytics for teacher training, while 2024 empirical research on ChatGPT-4 in higher education revealed positive student reception to real-time feedback despite sample size limitations [143]. The most significant technical advancement emerged in 2025 with Bi-LSTM networks augmented by attention mechanisms, enabling granular, interpretable MIDI performance analysis validated in piano pedagogy [33].

5 Discussion

5.1 Theoretical Understanding

Different models in creation intelligence exhibit distinct emotional mapping logics. For instance, GANs associate "sadness" with slow tempos and low registers, while VAEs emphasize minor harmonies. However, no unified standards exist for evaluating music; there is no consensus on how subjective emotional ratings relate to objective acoustic features. Beyond these emotional deviations, copyright and originality boundaries remain ambiguous. Models such as Jukebox and MusicLM, trained on vast copyrighted datasets, lack clear "similarity thresholds" to distinguish inspiration from infringement. A 2024 lawsuit over an AI-generated song sharing over 60% melodic similarity with existing works illustrates this issue. This reveals the broader human-AI responsibility debate and constraining AI creativity's commercialization.

Real-time alignment deviations persist in performance interactive systems. Even optimized platforms like MoMusic exhibit 100-200 ms latency during complex gestures (*e.g.*, rapid wrist rotations), disrupting expressive coherence. Meanwhile, debates over cultural authenticity persist in traditional instrument digitalization. AR bianzhong systems replicate timbres but fail to capture ancient performance techniques or temperament systems. These issues reflect the contradiction between 'technical replication' and 'cultural inheritance', as discussed under insufficient technical specificity in the introduction.

Most AI assessment tools are biased toward Western theory (*e.g.*, Kodály rhythm divisions, major/minor harmony rules), offering limited support for non-Western traditions such as erhu glissando or Indian raga modes. This leads to fairness biases in education intelligence, where non-Western techniques are misjudged as "technical errors". Moreover, the cognitive tension between "technical assistance" and "teacher replacement" persists. Although AI vocal training systems enable real-time intonation correction, emotional expression and aesthetic guidance still rely on human expertise. Nonetheless, some platforms overclaim, asserting that "AI can replace 80% of teachers' work", a claim that contradicts practice and fuels professional concern. This reflects the broader "gap between technical ideals and practical constraints", emphasizing that future AI music education should focus on human-AI collaboration rather than replacement.

5.2 Limitations and Gaps

Most current models are optimized for mainstream music styles (*e.g.*, Western classical and pop), with poor adaptability to non-mainstream styles such as Chinese folk music like guzheng and

bianzhong, African traditional rhythms, where [67] noted guzheng mode modeling requires separate optimization. Additionally, the same technology exhibits significant performance disparities across scenarios, *e.g.*, in music creation, short melody generation accuracy exceeds 85%, but structural coherence for extended musical structures, *e.g.*, movements longer than 5 minutes, remains a significant challenge. The other challenge is about the debates over human-AI responsibility division in artistic ethics – uncertainty regarding copyright ownership of AI-generated music, *e.g.*, whether rights belong to developers, users, or the AI itself for Jukebox-generated works, and the impact of AI performance on the "expressive uniqueness" of human artists, *e.g.*, defining artistic value between virtual violinists and human performers – most studies avoid in-depth discussions of cognitive and legal dimensions. A gap between technical ideals and practical constraints in real-world application – latency in multimodal performance systems and high scaling costs of AI education platforms (*e.g.*, over 10,000 RMB per VR music classroom device). While some studies, *e.g.*, [122] on ADMI collaboration costs, mention these issues, no systematic solutions exist, limiting technology transfer from laboratories to practical scenarios.

5.3 Ethical Concerns

As AI becomes increasingly integrated into music creation, the evolving relationship between human creators and intelligent systems raises complex ethical questions surrounding creative autonomy, algorithmic transparency, and cultural bias. [144] highlight musicians' ambivalence toward AI tools: while these systems assist in melody generation, audio mixing, and inspiration, they also risk eroding artistic identity and agency. Such tensions are intensified by the opaque "black box" nature of many generative models [145], which complicates questions of authorship, credit, and ownership. Beyond creative control, emerging research examines AI's engagement with the moral and cultural dimensions of music. Drawing on Moral Foundations Theory, [145] used GPT-4 and fine-tuned BERT to identify expressions of justice, care, and loyalty in song lyrics; this demonstrates AI's capacity to interpret moral content but also exposing risks of cultural bias when models trained on Western-centric data misrepresent non-Western traditions. These concerns mirror a broader issue across music creation, performance, and education: the gap between technical optimization and human-centered evaluation. In composition, the absence of perceptually grounded, culturally inclusive metrics limits assessment of AI-generated music's emotional and aesthetic value. Performance systems such as MoMusic [9] often privilege technical precision over expressive nuance, while educational platforms grounded in Western theory risk marginalizing diverse pedagogical practices.

6 Conclusion

Based on comprehensive bibliometric and systematic analyses, this study concludes that AI has profoundly transformed the music ecosystem across creation, performance, and education. Research output has surged by 14.92% annually since 2019, led by China, the United States, and the United Kingdom, which together contribute over half of global publications. The field demonstrates increasing interdisciplinary convergence as technological innovation, from rule-based systems to multimodal architectures such as Transformer-based MIDI-GPT and diffusion models, intersects with artistic expression, affective computing, and pedagogy, lowering creative barriers while enriching expressive depth and stylistic fidelity. AI-driven tools have redefined performance and learning: VR and AR systems enable gesture-based improvisation and immersive cultural re-enactments, while personalized and multimodal educational platforms enhance engagement and accessibility. Yet, critical challenges persist, including emotional authenticity, copyright ambiguity, real-time latency, cultural bias, and ethical concerns surrounding creative sovereignty and pedagogical integration. Future research should prioritize context-aware adaptive systems, explainable and fair AI design, cross-modal semantic alignment for accessibility, and clinically validated therapeutic applications. Sustained interdisciplinary collaboration will be vital to cultivating a globally inclusive ecosystem where technology augments (rather than replaces) musical creativity and education.

Ackownledgement

This work was supported by the Horizontal Research Project (NO. 39790102) from Ludong University and the Industrial R&D Project from DeepCox.

References

[1] R. Ramirez, E. Maestre, X. Serra, A rule-based evolutionary approach to music performance modeling, IEEE Transactions on Evolutionary Computation, 16 (2011) 96-107.

[2] B. Bozhanov, Computoser-rule-based, probability-driven algorithmic music composition, arXiv preprint arXiv:1412.3079, 2014.

[3] K. Hastuti, A. Azhari, A. Musdholifah, R. Supanggah, Rule-based and genetic algorithm for automatic gamelan music composition, International Review on Modelling and Simulations, 10 (2017) 202-212.

[4] Y. Chen, Y. Sun, The Usage of Artificial Intelligence Technology in Music Education System under Deep Learning, IEEE Access, 2024.

[5] Y. Chen, Innovation of Music Teaching Methods in Universities Based on Fuzzy Decision Support Systems and Deep Learning, International Journal of Fuzzy Systems, 2025.

[6] P. Suthaphan, V. Boonrod, N. Kumyaito, K. Tamee, Music generator for elderly using deep learning, In: Joint International conference on digital arts, media and technology with ECTI northern section conference on electrical, electronics, computer and telecommunication engineering, 2021, 289-292.

[7] J.-P. Briot, G. Hadjeres, F.-D. Pachet, Deep learning techniques for music generation, Springer, 2020.

[8] A. Huang, R. Wu, Deep learning for music, arXiv preprint arXiv:1606.04930, 2016.

[9] W. Bian, Y. Song, N. Gu, T.Y. Chan, T.T. Lo, T.S. Li, K.C. Wong, W. Xue, R.A. Trillo, MoMusic: A motion-driven human-AI collaborative music composition and performing system, In: AAAI Conference on Artificial Intelligence (AAAI), 37 (2023) 16057-16062.

[10] Z. Xu, D. Dutta, Y.-L. Wei, R.R. Choudhury, Multi-Source Music Generation with Latent Diffusion, arXiv preprint, 2025.

[11] W.-H. Kai, K.-X. Xing, Video-driven musical composition using large language model with memory-augmented state space, The Visual Computer, 2024, 1-13.

[12] P. Pasquier, J. Ens, N. Fradet, P. Triana, D. Rizzotti, J.-B. Rolland, M. Safi, MIDI-GPT: A Controllable Generative Model for Computer-Assisted Multitrack Music Composition, arXiv preprint arXiv:2501.17011, 2025.

[13] M. Grachten, S. Lattner, E. Deruty, Bassnet: A variational gated autoencoder for conditional generation of bass guitar tracks with learned interactive control, Applied Sciences, 10 (2020) 6627.

[14] S. Oore, I. Simon, S. Dieleman, D. Eck, K. Simonyan, This time with feeling: Learning expressive musical performance, Neural Computing and Applications, 32 (2020) 955-967.

[15] D. Kim, H.-W. Dong, D. Jeong, ViolinDiff: Enhancing Expressive Violin Synthesis with Pitch Bend Conditioning, arXiv preprint arXiv:2409.12477, 2024.

[16] Y.-J. Lin, H.-K. Kao, Y.-C. Tseng, M. Tsai, L. Su, A human-computer duet system for music performance, In: ACM International Conference on Multimedia, 2020, 772-780.

[17] D. Stefani, M. Tomasetti, F. Angeloni, L. Turchet, et al., Esteso: Interactive AI Music Duet Based on Player-Idiosyncratic Extended Double Bass Techniques, In: Proceedings of the International Conference on New Interfaces for Musical Expression (NIME'24), 2024.

[18] M. Sanganeria, R. Gala, Tuning Music Education: AI-Powered Personalization in Learning Music, arXiv preprint arXiv:2412.13514, 2024.

[19] Z. Ying, Experience of intelligent speech robot in music online classroom based on deep learning and virtual reality, Entertainment Computing, 52 (2025) 100795.

[20] J. Wu, X. Liu, X. Hu, J. Zhu, PopMNet: Generating structured pop music melodies using neural networks, Artificial Intelligence, 286 (2020) 103303.

[21] J. Grekow, T. Dimitrova, Monophonic music generation with a given emotion using conditional variational autoencoder, IEEE Access, 9 (2021) 129088-129101.

[22] W. Wang, J. Li, Y. Li, X. Xing, Style-conditioned Music Generation with Transformer-GANs, Frontiers of Information Technology & Electronic Engineering, 2024.

[23] K. Choi, J. Park, W. Heo, S. Jeon, J. Park, Chord conditioned melody generation with transformer-based decoders, IEEE Access, 9 (2021) 42071-42080.

[24] S. Lattner, M. Grachten, G. Widmer, Imposing higher-level structure in polyphonic music generation using convolutional restricted boltzmann machines and constraints, Journal of Creative Music Systems, 2 (2018) 1-31.

[25] X. Zhou, P. Yu, Social robots based on sensor technology simulate user music interaction experience, Entertainment Computing, 51 (2024) 100751.

[26] D. Wang, X. Guo, Research on evaluation model of music education informatization system based on machine learning, Scientific Programming, 2022.

[27] C. Hernandez-Olivan, J.R. Beltran, Music composition with deep learning: A review, In: Advances in Speech and Music Technology: Computational Aspects and Applications, 2022.

[28] C.-H. Liu, C.-K. Ting, Computational intelligence in music composition: A survey, IEEE Transactions on Emerging Topics in Computational Intelligence, 1 (2016) 2-15.

[29] M. Evin, A review on AI-enabled techniques for evaluating musician's performance, In: AIP Conference Proceedings, 3149 (2024).

[30] S. Holland, Artificial intelligence in music education: A critical review, In: Readings in Music and Artificial Intelligence, 2013, 239-274.

[31] J.F. Merchán Sánchez-Jara, S. González Gutiérrez, J. Cruz Rodríguez, B. Syroyid Syroyid, Artificial Intelligence-Assisted Music Education: A Critical Synthesis of Challenges and Opportunities, Education Sciences, 14 (2024) 1171.

[32] K. O'shea, R. Nash, An introduction to convolutional neural networks, arXiv preprint arXiv:1511.08458, 2015.

[33] Y. Han, Exploring a digital music teaching model integrated with recurrent neural networks under artificial intelligence, Scientific Reports, 2025.

[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is All you Need, Advances in Neural Information Processing Systems, 2017.

[35] P.L. Diéguez, V.-W. Soo, Variational autoencoders for polyphonic music interpolation, In: International Conference on Technologies and Applications of Artificial Intelligence (TAAI), 2020, 56-61.

[36] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Communications of the ACM, 63 (2020) 139-144.

[37] J. Zhang, G. Fazekas, C. Saitis, Composer Style-specific Symbolic Music Generation Using Vector Quantized Discrete Diffusion Models, arXiv preprint, 2024.

[38] P. Xiao, Enhancing emotional expression in algorithmic music composition systems using reinforcement learning, Journal of Computational Methods in Sciences and Engineering, 2025.

[39] L. Liu, R. Gong, Y. Yang, MusDiff: A multimodal-guided framework for music generation, Alexandria Engineering Journal, 129 (2025) 128-136.

[40] A. Agostinelli, T.I. Denk, Z. Borsos, J. Engel, M. Verzetti, A. Caillon, Q. Huang, A. Jansen, A. Roberts, M. Tagliasacchi, et al., Musiclm: Generating music from text, arXiv preprint arXiv:2301.11325, 2023.

[41] P. Dhariwal, H. Jun, C. Payne, J.W. Kim, A. Radford, I. Sutskever, Jukebox: A Generative Model for Music, arXiv preprint arXiv:2005.00341, 2020.

[42] C. Payne, MuseNet, OpenAI Blog, 3 (2019).

[43] Google AI, Magenta: Make Music and Art Using Machine Learning, `https://magenta.withgoogle.com/`, 2025.

[44] Aiva Technologies SARL, Personal AI music generation assistant, `https://www.aiva.ai`, 2025.

[45] S. Forsgren, H. Martiros, Riffusion - Stable diffusion for real-time music generation, `https://riffusion.com/about`, 2022.

[46] M. Aria, C. Cuccurullo, Bibliometrix: An R-tool for comprehensive science mapping analysis, Journal of Informetrics, 11 (2017) 959-975.

[47] C. Jin, T. Wang, X. Li, C.J.J. Tie, Y. Tie, S. Liu, M. Yan, Y. Li, J. Wang, S. Huang, A Transformer Generative Adversarial Network for Multi-track Music Generation, CAAI Transactions on Intelligence Technology, 2022.

[48] C. Gao, F. Reuben, T. Collins, Variation Transformer: New datasets, models, and comparative evaluation for symbolic music variation generation, In: Proceedings of the conference, 2024.

[49] J. Zhang, G. Fazekas, C. Saitis, Fast Diffusion GAN Model for Symbolic Music Generation Controlled by Emotions, arXiv preprint, 2023.

[50] C. Palmer, Music performance, Annual Review of Psychology, 1997.

[51] D.J. Hargreaves, N.A. Marshall, A.C. North, Music education in the twenty-first century: A psychological perspective, British Journal of Music Education, 2003.

[52] N. Van Eck, L. Waltman, Software survey: VOSviewer, a computer program for bibliometric mapping, Scientometrics, 2009.

[53] A. Ycart, E. Benetos, Learning and evaluation methodologies for polyphonic music sequence prediction with LSTMs, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28 (2020) 1328-1341.

[54] Y. Yan, Z. Duan, Measure by Measure: Measure-Based Automatic Music Composition with Modern Staff Notation, Transactions of the International Society for Music Information Retrieval (ISMIR), 2024.

[55] M. Tomasetti, L. Turchet, Handheld controller-based locomotion in Virtual Reality as an approach to interactive music composition: insights from composers' preferences, Digital Creativity, 2024.

[56] S. Zhang, X. Lu, X. Liu, Study on the Influence of AI Composition Software on Students' Creative Ability in Music Education, Journal of Educational Technology and Innovation (JETI), 2024.

[57] H. Pu, F. Jiang, Z. Chen, X. Song, ComposeOn Academy: Transforming Melodic Ideas into Complete Compositions Integrating Music Learning, arXiv preprint arXiv:2502.15255, 2025.

[58] M. Navarro-Caceres, H.G. Oliveira, P. Martins, A. Cardoso, Integration of a music generator and a song lyrics generator to create Spanish popular songs, Journal of Ambient Intelligence and Humanized Computing, 11 (2020) 4421-4437.

[59] X. Ma, Y. Wang, M.-Y. Kan, W.S. Lee, AI-lyricist: Generating music and vocabulary constrained lyrics, In: ACM International Conference on Multimedia, 2021.

[60] O. Vechtomova, G. Sahu, D. Kumar, Lyricjam: A system for generating lyrics for live instrumental music, arXiv preprint arXiv:2106.01960, 2021.

[61] X. Ma, V. Sharma, M.-Y. Kan, W.S. Lee, Y. Wang, KeYric: Unsupervised Keywords Extraction and Expansion from Music for Coherent Lyrics Generation, ACM Transactions on Multimedia Computing, Communications and Applications, 2024, 1-28.

[62] D.P. Kingma, M. Welling, et al., Auto-encoding variational bayes, In: International Conference on Learning Representations, 2013.

[63] A. Telikani, A. Tahmassebi, W. Banzhaf, A.H. Gandomi, Evolutionary Machine Learning: A Survey, ACM Computing Surveys, 54 (2022) 1-35.

[64] S. Tian, C. Zhang, W. Yuan, W. Tan, W. Zhu, XMusic: Towards a Generalized and Controllable Symbolic Music Generation Framework, arXiv preprint, 2025.

[65] A. Muhamed, L. Li, X. Shi, S. Yaddanapudi, W. Chi, D. Jackson, R. Suresh, Z.C. Lipton, A.J. Smola, Symbolic music generation with transformer-gans, In: AAAI Conference on Artificial Intelligence (AAAI), 35 (2021) 408-417.

[66] P. Neves, J. Fornari, J. Florindo, Generating Music with Sentiment using Transformer-GANs, arXiv preprint, 2022.

[67] Y.-H. Lan, W.-Y. Hsiao, H.-C. Cheng, Y.-H. Yang, MusiConGen: Rhythm and Chord Control for Transformer-based Text-to-Music Generation, arXiv preprint, 2024.

[68] C.-Z.A. Huang, A. Vaswani, J. Uszkoreit, N. Shazeer, C. Hawthorne, A.M. Dai, M.D. Hoffman, D. Eck, An Improved Relative Self-Attention Mechanism for Transformer with Application to Music Generation, CoRR, abs/1809.04281 (2018).

[69] C. Donahue, H.H. Mao, Y.E. Li, G.W. Cottrell, J. McAuley, LakhNES: Improving multi-instrumental music generation with cross-domain pre-training, arXiv preprint arXiv:1907.04868, 2019.

[70] J. Ens, P. Pasquier, Mmm: Exploring conditional multi-track music generation with the transformer, arXiv preprint arXiv:2008.06048, 2020.

[71] Y.-S. Huang, Y.-H. Yang, Pop music transformer: Beat-based modeling and generation of expressive pop piano compositions, In: ACM International Conference on Multimedia, 2020, 1180-1188.

[72] G. Hadjeres, L. Crestel, The piano inpainting application, arXiv preprint arXiv:2107.05944, 2021.

[73] D. Makris, K.R. Agres, D. Herremans, Generating lead sheets with affect: A novel conditional seq2seq framework, In: International Joint Conference on Neural Networks (IJCNN), 2021, 1-8.

[74] J. Zhou, H. Zhu, X. Wang, Choir Transformer: Generating Polyphonic Music with Relative Attention on Transformer, arXiv preprint, 2023.

[75] J. Luo, X. Yang, D. Herremans, BandControlNet: Parallel Transformers-based Steerable Popular Music Generation with Fine-Grained Spatiotemporal Features, arXiv preprint, 2024.

[76] Y. Zhu, K. Olszewski, Y. Wu, P. Achlioptas, M. Chai, Y. Yan, S. Tulyakov, Quantized GAN for Complex Music Generation from Dance Videos, arXiv preprint, 2022.

[77] M. Han, S. Soradi-Zeid, T. Anwlnkom, Y. Yang, Firefly Algorithm-based LSTM Model for Guzheng Tunes Switching with Big Data Analysis, Heliyon, 2024.

[78] J. Jeong, Y. Kim, C.W. Ahn, A multi-objective evolutionary approach to automatic melody generation, Expert Systems with Applications (ESWA), 90 (2017) 50-61.

[79] H.B. Lopes, F.V.C. Martins, R.T.N. Cardoso, V.F. dos Santos, Combining rules and proportions: A multiobjective approach to algorithmic composition, In: IEEE Congress on Evolutionary Computation (CEC), 2021, 2282-2289.

[80] C. De Felice, R. De Prisco, D. Malandrino, G. Zaccagnino, R. Zaccagnino, R. Zizza, Splicing music composition, Information Sciences, 385 (2017) 196-212.

[81] N. Masuda, H. Iba, Musical composition by interactive evolutionary computation and latent space modeling, In: IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2018, 2792-2797.

[82] F. Mo, X. Wang, S. Li, H. Qian, A music generation model for robotic composers, In: IEEE International Conference on Robotics and Biomimetics (ROBIO), 2020, 1483-1488.

[83] B. Stoltz, A. Aravind, MU_PSYC: music psychology enriched genetic algorithm, In: IEEE Congress on Evolutionary Computation (CEC), 2019, 2121-2128.

[84] Y.-W. Wen, C.-K. Ting, Composing bossa nova by evolutionary computation, In: International Joint Conference on Neural Networks (IJCNN), 2020, 1-8.

[85] N. Shi, Y. Wang, Symmetry in computer-aided music composition system with social network analysis and artificial neural network methods, Journal of Ambient Intelligence and Humanized Computing, 2020, 1-16.

[86] R. De Prisco, G. Zaccagnino, R. Zaccagnino, Evo-Composer: An evolutionary algorithm for 4-voice music compositions, Evolutionary Computation, 28 (2020) 489-530.

[87] R. Sabitha, S. Majji, M. Kathiravan, S.G. Kumar, K.G. Kharade, S.R. Karanam, Artificial intelligence based music composition system - multi-algorithmic music arranger, In: International Conference on Electronics and Sustainable Communication Systems (ICESC), 2021, 1808-1813.

[88] Z. Zeng, L. Zhou, A memetic algorithm for Chinese traditional music composition, In: International Conference on Intelligent Computing and Signal Processing (ICSP), 2021, 187-192.

[89] L.R. De Azevedo Santos, C.N. Silla Jr, M.D. Costa-Abreu, A methodology for procedural piano music composition with mood templates using genetic algorithms, In: International Conference of Pattern Recognition Systems (ICPRS), 2021, 1-6.

[90] J. Kilb, C. Ellis, Conserving Human Creativity with Evolutionary Generative Algorithms: A Case Study in Music Generation, arXiv preprint arXiv:2406.05873, 2024.

[91] Z. Qiu, Y. Ren, C. Li, H. Liu, Y. Huang, Y. Yang, S. Wu, H. Zheng, J. Ji, J. Yu, et al., Mind band: a cross-media AI music composing platform, In: ACM International Conference on Multimedia, 2019, 2231-2233.

[92] P.-S. Cheng, C.-Y. Lai, C.-C. Chang, S.-F. Chiou, Y.-C. Yang, A variant model of TGAN for music generation, In: Asia Service Sciences and Software Engineering Conference (ASSE), 2020, 40-45.

[93] T. Wang, J. Liu, C. Jin, J. Li, S. Ma, An intelligent music generation based on Variational Autoencoder, In: International Conference on Culture-oriented Science & Technology (ICCST), 2020, 394-398.

[94] C.-F. Huang, C.-Y. Huang, Emotion-based AI music generation system with CVAE-GAN, In: IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), 2020, 220-222.

[95] M.W.Y. Lam, Q. Tian, T. Li, Z. Yin, S. Feng, M. Tu, Y. Ji, R. Xia, M. Ma, X. Song, J. Chen, Y. Wang, Y. Wang, Efficient Neural Music Generation, arXiv preprint arXiv:2305.15719, 2023.

[96] Y. Gan, Attention-Guided Music Generation with Variational Autoencoder and Latent Diffusion, In: International Workshop on Materials Engineering and Computer Sciences (IWMECS), 2024.

[97] A. Tikhonov, I.P. Yamshchikov, et al., Music generation with variational recurrent autoencoder supported by history, arXiv preprint arXiv:1705.05458, 2017.

[98] G. Brunner, A. Konrad, Y. Wang, R. Wattenhofer, MIDI-VAE: Modeling Dynamics and Instrumentation of Music with Applications to Style Transfer, Computing Research Repository CoRR, 2018.

[99] A. Roberts, J. Engel, C. Raffel, C. Hawthorne, D. Eck, A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music, In: International Conference on Machine Learning (ICML), 2018.

[100] S. Lattner, M. Grachten, High-Level Control of Drum Track Generation Using Learned Patterns of Rhythmic Interaction, In: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2019.

[101] H.-T. Hung, C.-Y. Wang, Y.-H. Yang, H.-M. Wang, Improving Automatic Jazz Melody Generation by Transfer Learning Techniques, In: Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2019.

[102] B. Jia, J. Lv, Y. Pu, X. Yang, Impromptu accompaniment of pop music using coupled latent variable model with binary regularizer, In: International Joint Conference on Neural Networks (IJCNN), 2019, 1-6.

[103] Y.-Q. Lim, C.S. Chan, F.Y. Loo, ClaviNet: Generate music with different musical styles, IEEE MultiMedia, 28 (2020) 83-93.

[104] C. Jin, Y. Tie, Y. Bai, X. Lv, S. Liu, A style-specific music composition neural network, Neural Processing Letters, 52 (2020) 1893-1912.

[105] J. Zhu, K. Sakurai, R. Togo, T. Ogawa, M. Haseyama, MMT-BERT: Chord-aware Symbolic Music Generation Based on Multitrack Music Transformer and MusicBERT, arXiv preprint, 2024.

[106] R. Manzelli, V. Thakkar, A. Siahkamari, B. Kulis, Conditioning deep generative raw audio models for structured automatic music, arXiv preprint arXiv:1806.09905, 2018.

[107] P. Wiriyachaiporn, K. Chanasit, A. Suchato, P. Punyabukkana, E. Chuangsuwanich, Algorithmic music composition comparison, In: International Joint Conference on Computer Science and Software Engineering (JCSSE), 2018, 1-6.

[108] N. dos Santos Cunha, A. Subramanian, D. Herremans, Generating guitar solos by integer programming, Journal of the Operational Research Society, 69 (2018) 971-985.

[109] Y. Huang, A. Ghatare, Y. Liu, Z. Hu, Q. Zhang, C.S. Sastry, S. Gururani, S. Oore, Y. Yue, Symbolic Music Generation with Non-Differentiable Rule Guided Diffusion, arXiv preprint, 2024.

[110] Soundraw, Create custom music and beats with AI, https://soundraw.io, 2025.

[111] ShutterstockInc., Amper Music, https://www.shutterstock.com/discover/amper-music, 2025.

[112] Ecrett Music, Royalty Free Music for Creators, https://ecrettmusic.com, 2018.

[113] Boomy Corporation, Boomy, https://boomy.com, 2023.

[114] AI Tool Selection, Discover AI Tools for Your Daily Tasks, https://aitoolselection.com/zh-CN, 2025.

[115] H.-W. Dong, W.-Y. Hsiao, L.-C. Yang, Y.-H. Yang, Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment, In: AAAI Conference on Artificial Intelligence (AAAI), 32 (2018).

[116] K. Kumar, R. Kumar, T. de Boissiere, L. Gestin, W.Z. Teoh, J. Sotelo, A. de Brebisson, Y. Bengio, A. Courville, MelGAN: Generative Adversarial Networks for Conditional Waveform Synthesis, arXiv preprint, 2019.

[117] J. Engel, L. (Hanoi) Hantrakul, C. Gu, A. Roberts, DDSP: Differentiable Digital Signal Processing, In: International Conference on Learning Representations, 2020.

[118] R. Yuan, H. Lin, S. Guo, G. Zhang, J. Pan, et al., YuE: Scaling Open Foundation Models for Long-Form Music Generation, arXiv preprint, 2025.

[119] H. Zulić, How AI can change/improve/influence music composition, performance and education: three case studies, INSAM Journal of Contemporary Music, Art and Technology, 2019, 100-114.

[120] S. Ppali, V. Lalioti, B. Branch, C.S. Ang, A.J. Thomas, B.S. Wohl, A. Covaci, Keep the VRhythm going: A musician-centred study investigating how Virtual Reality can support creative musical practice, In: CHI Conference on Human Factors in Computing Systems, 2022.

[121] W. Guo, Y. Huang, Z. Chen, Z. Zhang, G. Sun, Q. Zeng, X. Li, The “rebirth” of traditional musical instrument: An interactive installation based on augmented reality and somatosensory technology to empower the exhibition of chimes, Computer Animation and Virtual Worlds, 2023.

[122] H. Lindetorp, M. Svahn, J. Hölling, K. Falkenberg, E. Frid, Collaborative music-making: special educational needs school assistants as facilitators in performances with accessible digital musical instruments, Frontiers in Computer Science, 5 (2023).

[123] Y. Ma, C. Wang, Empowering music education with technology: a bibliometric perspective, Humanities and Social Sciences Communications, 2025.

[124] M. Biasutti, Strategies adopted during collaborative online music composition, International Journal of Music Education, 2018.

[125] X. Wang, Design of vocal music teaching system platform for music majors based on artificial intelligence, Wireless Communications and Mobile Computing, 2022.

[126] M.C. Angelides, A.K.Y. Tong, Implementing multiple tutoring strategies in an intelligent tutoring system for music learning, Journal of Information Technology, 10 (1995) 52-62.

[127] A. Ara, R. Gopalakrishna, A Study on Emotion Identification from Music Lyrics, In: International Conference of Reliable Information and Communication Technology (IRICT), 2020, 396-406.

[128] Y. Shi, Exploring Music Teaching Methods Through Core Literacy: A Deep Learning Approach with Implications for Cognitive and Emotional Development in Sports, Revista de Psicología del Deporte (Journal of Sport Psychology), 2023.

[129] F. Sun, Analysis of Virtual Reality-based Music Education Experience and its Impact on Learning Outcomes, Scalable Computing: Practice and Experience, 25 (2024) 4755-4762.

[130] K. Han, W. You, S. Shi, L. Sun, Hearing with the eyes: modulating lyrics typography for music visualization, The Visual Computer, 2024.

[131] A. Vargas, P. Díaz, T. Zarraonandia, Using virtual reality and music in cognitive disability therapy, In: International Conference on Advanced Visual Interfaces (AVI), 2020, 1-9.

[132] P. Pérez, M. Orduna, M. Nava-Ruiz, J. Martín-Boix, Using immersive video to recall significant musical experiences in elderly population with intellectual disability, In: IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), 2024, 887-888.

[133] E. Aras, Style learning and musical mimicry in Artificial Intelligence: modern approaches, Journal of AI, Humanities and New Ethics, 2025, 19-32.

[134] J. Fang, Artificial intelligence robots based on machine learning and visual algorithms for interactive experience assistance in music classrooms, Entertainment Computing, 52 (2025) 100779.

[135] D.T. Larose, C.D. Larose, K-nearest Neighbor Algorithm, 2014.

[136] C. Lee, J.-H. Hong, musicolors: Bridging Sound and Visuals For Synesthetic Creative Musical Experience, arXiv preprint arXiv:2503.14220, 2025.

[137] A. Dash, K. Agres, AI-based affective music generation systems: A review of methods and challenges, ACM Computing Surveys, 56 (2024) 1-34.

[138] L. Schaab, A. Kruspe, Joint sentiment analysis of lyrics and audio in music, arXiv preprint arXiv:2405.01988, 2024.

[139] J. Tobolewski, M. Sakowicz, J. Turmo Borras, B. Kostek, A bimodal deep model to capture emotions from music tracks, Journal of artificial intelligence and soft computing research, 15 (2025) 215-238.

[140] E.G. Duarte, I. Cossette, M.M. Wanderley, Analysis of Accessible Digital Musical Instruments through the lens of disability models: a case study with instruments targeting d/Deaf people, Frontiers in Computer Science, 5 (2023) 1158476.

[141] M. Biasutti, Assessing a collaborative online environment for music composition, Journal of Educational Technology & Society, 18 (2015) 49-63.

[142] Z.H. Yun, Y. Alshehri, N. Alnazzawi, I. Ullah, S. Noor, N. Gohar, A decision-support system for assessing the function of machine learning and artificial intelligence in music education for network games, Soft Computing, 2022.

[143] J. Xi, Artificial Intelligence Technology in the Assessment of Teachers' Music Teaching Skills Training, International Journal of Educational Innovation and Science, 2023.

[144] M. Newman, L. Morris, J.H. Lee, Human-AI Music Creation: Understanding the Perceptions and Experiences of Music Creators for Ethical and Productive Collaboration., In: International Society for Music Information Retrieval (ISMIR), 2023.

[145] V. Preniqi, I. Ghinassi, J. Ive, K. Kalimeri, C. Saitis, Automatic Detection of Moral Values in Music Lyrics, arXiv preprint arXiv:2407.18787, 2024.

Fei Tong is a violinist and nationally certified teacher of music (NCTM), currently serving as an associate professor in violin at Ludong University, China. She has performed internationally across Asia, North America, and Europe, appearing as a soloist with the Harbin Symphony Orchestra, University of Georgia Symphony Orchestra, and ARCO Chamber Orchestra, and performing at prestigious venues including Carnegie Hall, the Kennedy Center, Royal Albert Hall, and Beijing's National Center for the Performing Arts. Prof Tong earned her Doctor of Musical Arts (DMA) in violin performance from the University of Georgia (USA) and studied at the China Conservatory of Music. Her honors include being the National Winner of The American Prize 2025 and top prizes at competitions such as the Alexander & Buono International String Competition and London Grand Prize Virtuoso International Music Competition. Prof Tong has published several research papers, authored one monograph, and developed one instructional textbook for violin education.
https://orcid.org/0000-0001-6194-5823

Dongjing Jiang is an AI research assistant at DeepCox Intelligence and a Guzheng (Chinesee plucked zither) performance master. She has participated in the music intelligence program at OKIC, combining her expertise in music and technology. Her research focuses on music intelligence, data mining, and the development of digital music tools.

Qingchong Jiao is currently a stack manager and data scientist at DeepCox Intelligence. He received his B.Eng. degree in Data Science and Big Data Technology from Hebei University of Environmental Engineering, China. His research interests include artificial intelligence, machine learning, data mining, and computer vision.

Albina Isufi is a Kosovo-born soprano, voice and piano instructor at the Music School of Westchester (USA), and an advisor at DeepCox (Canada). She holds Bachelor's and Master's degrees in Opera (Vocal Performance) from the Stockholm University of the Arts (Sweden) and completed advanced operatic training at Operastudio in Stockholm. She has performed extensively across Europe in operatic and concert repertoire, with appearances at leading institutions including Royal Opera, Norrlandsoperan, Confidencen, Folkoperan, and Musiikkitalo. Her operatic roles include Fiordiligi (Così fan tutte), Contessa (Le nozze di Figaro), Euridice (Orfeo ed Euridice), and Liu (Turandot).

Flynnwell Jianfei Zhang serves as Director of OKIC (Canada) and Principal of DeepCox Intelligence (Canada). He earned his PhD in Artificial Intelligence from Université de Sherbrooke, Canada in 2019, followed by postdoctoral research at Case Western Reserve University, USA (2020) and University of Alberta / Amii, Canada (2021). Dr. Zhang has published over 20 research papers and, as the first author, received the ACM CIKM 2021 Best Paper Award (Full Paper) and the PAKDD 2019 Best Application Paper Award. His research focuses on knowledge intelligence and interdisciplinary AI applications.
https://orcid.org/0000-0002-6303-3390