Human–AI Co-Creation for Accessible Music Education and Inclusive Creative Experiences

Emile Makinen
Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
makinenemile@buffalo.edu

Leon D. Lane
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
leonmail@unr.edu

Enzo Hughes
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.
hughes698@oregonstate.edu

Jeffrey Bergman
Department of Computer Science, George Mason University, Fairfax, VA, USA.
bergman1982@gmu.edu

Abstract

Advances in generative and interactive artificial intelligence have opened new possibilities for music education, yet translating these technical capabilities into accessible and inclusive learning remains a systems-level challenge. This article examines human-AI co-creation as a sociotechnical infrastructure for music education, focusing on structural trade-offs among pedagogical agency, algorithmic mediation, data governance, institutional readiness, and long-term equity. Drawing on scholarship in human-computer interaction, AI in education, music information retrieval, fairness, and platform governance, the article argues that accessible co-creation systems cannot be reduced to improved model performance or novel interface design. They require modular architectures that separate sensing, musical interpretation, and feedback while preserving low-latency interaction for learners with diverse sensory, cognitive, and motor profiles. They also require governance arrangements that address representational bias in training corpora, sustainable deployment pathways for under-resourced institutions, and transparent evaluation practices aligned with musical learning rather than narrow task accuracy. The discussion analyzes conceptual frameworks and system-level cases to show how co-creative platforms can shift power relations among learners, educators, institutions, and technology providers. The article concludes with policy implications for procurement, interoperability, privacy, and professional development, and outlines a research agenda centered on fairness, robustness, and inclusive musical participation. The intended audience includes system designers, educators, and policymakers concerned with building accessible creative infrastructures that remain responsive to human values.

Keywords

human-AI co-creation; accessible music education; inclusive creative experiences; sociotechnical systems; algorithmic fairness; educational infrastructure.

1. Introduction

Music education has historically been shaped by access to instruments, notation literacy, individualized instruction, and institutional resources. These conditions create persistent barriers for learners with disabilities, learners in low-income communities, and learners in remote regions. Human-AI co-creation systems promise alternative pathways by translating diverse gestures, sounds, and preferences into musical structures, generating adaptive accompaniment, and offering continuous feedback. However, introducing AI into creative learning is not merely a technical addition. It alters roles, expectations, and pedagogical relationships. Research on human-AI interaction emphasizes that systems should communicate capabilities and limitations clearly, support efficient correction, and maintain user control [1]. Similarly, AI literacy scholarship argues that learners and educators must be able to understand, interrogate, and shape AI behavior rather than being passive recipients of algorithmic outputs [2].

The particular domain of music adds sensory and temporal constraints that distinguish it from text-based educational technologies. Longstanding work on digital musical instruments shows that expressive control depends on mappings between human action and sonic response, and that accessibility can be improved when interfaces move beyond conventional keyboard or staff notation assumptions [3]. Interactive machine learning systems have further demonstrated that real-time model training can allow users to define their own gesture-sound relationships, thereby supporting forms of musical expression that would otherwise be difficult to encode in advance [4]. These developments suggest that accessible music education platforms should be understood as neither purely educational software nor purely musical instruments, but as hybrid systems that must negotiate real-time interaction, pedagogical scaffolding, creative agency, and institutional constraints.

This article develops a systems-level perspective on human-AI co-creation for accessible music education. It examines the structural trade-offs involved in designing, deploying, and governing such systems. Rather than proposing a single technical architecture or pedagogical method, the article analyzes how architectural choices, data practices, interface design, institutional deployment, evaluation, and policy interact to produce or undermine inclusive creative experiences. The discussion emphasizes that accessibility and inclusion are emergent properties of the entire sociotechnical arrangement, not features that can be added after a model is trained or an interface is built.

2. Human-AI Co-Creation as a Sociotechnical System

Human-AI co-creation in music education involves learners, teachers, institutions, curricula, datasets, models, interfaces, and platform operators. A change in any one of these elements can reshape the others. Early critical reviews in music and AI cautioned against treating computational tutors as neutral authorities because pedagogical goals and musical values are embedded in system design [5]. The same caution applies to contemporary co-creative platforms. A system that emphasizes chord substitution, for example, may implicitly privilege harmonic conventions from particular musical traditions. A system that rewards pitch accuracy may marginalize timbral or rhythmic creativity. These are not neutral engineering decisions; they reflect assumptions about what counts as musical learning and whose expressive goals matter.

Arguments for AI in education frequently emphasize personalization, but they also depend on institutional data infrastructures, teacher training, and pedagogical integration [6]. Personalized learning cannot be delivered simply by making a model adaptive. It requires sustained interaction between the learner, the educator, and the system, as well as institutional capacity to manage data, devices, curriculum alignment, and professional development. Holmes and colleagues describe AI in education as a set of competing design philosophies, including learner-centered, teacher-centered, and system-centered approaches [7]. In music education, this competition is especially visible. A learner-centered system might allow open exploration and user-defined mappings. A teacher-centered system might prioritize curriculum sequencing and assessment. A system-centered design might optimize for technical novelty or generative fluency. These orientations are not always compatible, and their relative priority must be negotiated through design and governance rather than assumed.

The ethical governance of such systems requires attention to beneficence, non-maleficence, autonomy, justice, and explicability [8]. In music education, justice includes fair representation of genres, languages, cultural practices, and disability communities within training data and evaluation procedures. Explicability includes helping teachers and learners understand why a system generated a particular accompaniment, suggested a modification, or failed to respond to a gesture. Autonomy requires that users can decline algorithmic suggestions without losing access to meaningful musical participation. These ethical dimensions are not separate from technical design. They become concrete in decisions about data collection, model complexity, interface defaults, and the distribution of control between human and machine.

3. Architectural Considerations for Accessible Music Learning Environments

A well-designed co-creative music education system generally requires a modular architecture that separates sensing, musical representation, generative decision-making, and feedback. Modularity allows components to be replaced or adapted for different learner needs without requiring a full system redesign. A sensing layer may detect motion, audio input, touch, gaze, or physiological signals. A musical representation layer may translate those signals into pitch, rhythm, dynamics, or timbral descriptors. A generative layer may produce accompaniment, harmonization, or responsive textures. A feedback layer may present visual, auditory, or haptic information to the learner. Each layer involves distinct technical challenges, yet their integration determines whether the system feels responsive, understandable, and controllable.

Latency is especially critical in music. Even small delays between a gesture and its sonic result can disrupt musical timing and reduce expressive agency. This creates a structural tension between cloud-based architectures and edge-based architectures. Cloud infrastructure can support larger generative models, frequent updates, and centralized analytics. However, it introduces network dependency, privacy risks, and variable latency that may be unacceptable in musical interaction. Edge deployment can reduce latency and keep sensitive learner data local, but it may increase hardware costs and complicate maintenance in under-resourced schools. Accessibility therefore depends not only on algorithmic design but also on infrastructural decisions about where computation occurs, who maintains devices, and how systems behave when connectivity is limited.

Interoperability is another architectural concern. Accessible music education platforms rarely operate in isolation. They must interact with learning management systems, assistive technologies, student information systems, and existing music software. Without common representation standards for musical gestures, scores, and audio events, platforms can become silos that fragment learner progress and increase the burden on educators. Open interfaces and documented data formats support portability and reduce vendor lock-in. At the same time, interoperability can expose learner data across multiple systems, increasing the need for consistent privacy controls and access policies. The architectural trade-off is therefore not simply between open and closed systems, but between interoperability, privacy, security, and institutional manageability.

4. Inclusive Creative Interaction and Interface Design

Accessible music education requires universal design rather than retrofitting after a system has been built. Multimodal interfaces can support learners with diverse sensory, motor, and cognitive profiles by combining visual notation, auditory feedback, haptic cues, and motion input. A central lesson from fairness research is that recognition systems may perform unevenly across demographic groups, and that such disparities often become visible only when evaluation moves beyond aggregate accuracy [9]. In a music co-creation platform, gesture recognition, voice processing, and pose estimation may fail disproportionately for users with atypical movement patterns, nonstandard vocal characteristics, or assistive device use. These failures are not merely technical bugs; they can exclude the very learners the system is intended to include.

Selbst and colleagues argue that fairness cannot be solved through isolated technical adjustments because social context shapes the meaning and impact of algorithmic systems [10]. This insight applies directly to inclusive interface design. A gesture-sound mapping that works for a seated learner using fine finger control may not work for a learner who communicates through eye movement or gross motor gestures. Training datasets that capture predominantly Western concert music movement patterns may not represent culturally specific musical gestures or disability-related motor patterns. Participatory design with disability communities and musicians from multiple traditions is therefore necessary to identify how algorithmic abstraction may erase relevant context or create new barriers.

Creative agency is a central design requirement. Accessible interfaces should allow users to define mappings, choose genres, modify complexity, and interrupt suggestions. This can conflict with the goal of reducing cognitive load. For a learner with limited motor control, a system that predicts the next chord may reduce physical effort, but it may also remove a meaningful creative decision. A co-creative system can address this tension by offering adjustable autonomy, in which the user controls the degree of suggestion, correction, and generation. Such designs require careful attention to feedback: the interface must communicate what the system is doing, what it expects, and how the learner can take control. These principles align with established guidance for human-AI interaction, which emphasizes user control, transparent feedback, and graceful recovery from error [1].

5. Data Governance, Bias, and Representation in Musical AI

Training data for musical AI are often skewed toward Western tonal music, common instruments, professional recordings, and copyright-cleared materials. This skew has both cultural and educational consequences. Crawford has argued that AI training data are not neutral raw material but encode histories of power, labor, and value [11]. When a generative music system is trained predominantly on a narrow repertoire, its outputs may treat other musical traditions as deviations or fail to recognize their structures. For learners in those traditions, the system may feel irrelevant or, worse, validate exclusion by presenting the dominant repertoire as the default for musical competence.

Eubanks demonstrates that algorithmic systems in public services, when deployed without accountable governance, can intensify existing inequalities rather than remedy them [12]. In educational contexts, predictive or adaptive systems may categorize learners based on historical patterns, limiting access to advanced creative opportunities. A music education platform that assigns simple rhythmic exercises to learners with certain demographic or disability characteristics may reproduce lower expectations. Similarly, Williamson describes how big data in education shifts authority toward platform providers and away from teachers and local institutions [13]. Data-driven music platforms can create new dependencies in which pedagogical decisions are shaped by vendor dashboards and proprietary metrics rather than by educators' situated knowledge.

Recent system-level experiments in multimodal musical co-creation beyond traditional classrooms illustrate both the therapeutic potential of such platforms and the governance questions they raise when affective and behavioral data are collected from vulnerable participants [14]. Although not primarily an educational intervention, this work highlights the breadth of data that co-creative systems can capture once multimodal sensing, generative music, and personal interaction are combined. It also underscores the need to integrate privacy protections, informed consent, and participant welfare into the design of creative systems that may operate across educational, therapeutic, and domestic settings.

Data governance for accessible music education must address informed consent, data minimization, cultural ownership, copyright, and benefit sharing. Audio recordings of students may reveal identity, disability status, emotional state, or household conditions. Gesture data may reveal mobility patterns. Because music is also cultural expression, the use of student performances to fine-tune generative models raises questions about who benefits from that creative labor. Cross-domain lessons from health and social services suggest that governance mechanisms should include independent review, local data stewardship, and clear accountability for algorithmic harms. In music education, these mechanisms are still underdeveloped, particularly where platforms are procured by schools and used by minors.

6. Institutional Deployment and Infrastructure Sustainability

Deployment of human-AI music co-creation systems in schools and community organizations requires integration with existing technical support structures. Many music classrooms have limited bandwidth, aging devices, and minimal access to specialized IT staff. Automatic music transcription and source separation systems can reduce the complexity of working with recorded music, but they require computational resources and may not function well with classroom noise or unconventional instruments [15]. A platform designed for high-end hardware may fail in the very environments where accessible music education is most needed. Sustainable deployment therefore demands explicit attention to minimum hardware requirements, offline functionality, and graceful degradation under constrained conditions.

Sturm and colleagues argue that machine learning research for music creation should be evaluated by musical outcomes and practitioner use, not only by quantitative benchmarks [16]. This argument has direct implications for institutional deployment. A system that performs well on a standard dataset may still be unusable in a classroom because it does not align with the teacher's pedagogical goals, the students' musical backgrounds, or the available instruments. Institutions should be able to pilot systems in realistic conditions and assess their value through a combination of technical performance, pedagogical fit, accessibility, and learner engagement. Procurement processes that rely solely on vendor demonstrations or aggregate accuracy metrics are unlikely to capture these dimensions.

Ubiquitous music research emphasizes ecologically grounded design that connects digital tools with local musical practices [17]. This perspective supports a shift from centralized platform thinking toward modular, community-configurable infrastructures. Local hubs can host services, manage data, and adapt interfaces to local languages and musical traditions. Such arrangements can reduce dependence on continuous internet connectivity and international data transfers. They also create opportunities for community ownership and maintenance. However, localized infrastructure introduces its own sustainability challenges, including the need for ongoing technical support, device repair, and educator training. Without those supports, decentralization can become another form of neglect.

Professional transformation is also central to sustainable deployment. Susskind and Susskind argue that professional work is being decomposed and redistributed through technology [18]. In music education, AI may take over routine transcription, ear-training drills, or accompaniment generation, freeing educators for more interpretive and relational teaching. Yet this redistribution is not automatic. It requires professional development that helps teachers understand system behavior, contest algorithmic suggestions, and integrate co-creative tools into meaningful musical activities. If teacher training lags behind platform deployment, AI may become a substitute for skilled instruction rather than a complement to it. The structural challenge is to deploy systems in ways that enhance educator capacity without intensifying workload or undermining professional judgment.

7. Robustness, Evaluation, and Pedagogical Alignment

Robustness in creative systems differs from robustness in classification or prediction tasks. Musical outputs are open-ended, context-dependent, and evaluated through aesthetic and pedagogical criteria that resist simple automation. Bender and colleagues caution that scale and fluency do not guarantee understanding, and that evaluation practices inherited from narrow tasks can obscure broader harms [19]. In music generation, fluency can produce plausible but pedagogically misleading harmonization or stylistically inconsistent improvisation. A system that always produces competent-sounding output may hide the fact that it does not understand the learner's intentions, the musical genre, or the pedagogical objective.

Evaluation therefore needs to address perceptual quality, pedagogical value, accessibility, cultural relevance, and learner agency. These dimensions cannot be collapsed into a single metric. A co-creative platform may generate high-quality audio yet reduce learner autonomy by making too many decisions. It may improve pitch accuracy on standardized exercises while discouraging improvisation. It may be robust across studio conditions but fail in noisy classrooms or with nonstandard instruments. Longitudinal evaluation is needed to understand how learners develop musical identity, creative confidence, and technical skill over time. Such evaluation should include disabled users, non-Western musical traditions, and learners with limited prior formal training.

Pedagogical alignment is also a robustness concern. Situated learning theory suggests that legitimate peripheral participation in musical communities is essential for developing identity and skill [20]. This implies that AI co-creation should support participation in musical communities, not isolate learners with screens and automated feedback. A platform that adapts only to individual performance may overlook the social dimensions of music learning, such as ensemble playing, listening, negotiation, and shared performance. Robust evaluation should therefore consider whether the system fosters or undermines social musical participation. This may include assessing opportunities for peer collaboration, teacher involvement, and connection to local musical cultures.

8. Policy Implications and Long-Term Governance

Public institutions should adopt procurement policies that require accessible design, interoperability, data protection, and independent evaluation for AI music education platforms. Such policies can prevent vendor lock-in and ensure that platforms remain adaptable to changing curricula and learner needs. Procurement should also consider total cost of ownership, including hardware, connectivity, maintenance, training, and data governance. A low-cost initial license may become expensive if it requires proprietary sensors, continuous cloud subscriptions, or frequent paid updates. Policy instruments can require vendors to document algorithmic behavior, data flows, and accessibility testing before deployment in schools.

Student privacy is a critical governance issue. Educational platforms are often governed by student data protection laws, but creative audio and gesture data may be more sensitive than conventional academic records. Recordings of singing, playing, or movement can reveal disability status, emotional state, and cultural background. Data protection policies should require informed consent, data minimization, retention limits, and clear deletion procedures. Schools should have the capacity to audit what data are collected and to contest automated decisions that affect learners. Cross-domain lessons from healthcare and social services suggest that independent oversight bodies can help ensure that creative AI systems do not reproduce discrimination or erode trust.

Equity requires attention to ongoing operational costs and community capacity. Open-source and community-owned models can support local adaptation, but they require sustained investment in maintenance and technical expertise. Public funding should therefore support not only initial adoption but also long-term infrastructure, accessible documentation, and professional learning communities. Algorithmic impact assessments should be conducted before large-scale deployment and updated as systems evolve. These assessments should include disabled learners, educators, caregivers, and musicians from marginalized traditions. Long-term governance also requires mechanisms for redress when systems fail, exclude, or cause distress. Without such mechanisms, the burden of adaptation falls disproportionately on the least powerful users.

The research agenda should include longitudinal studies of learning outcomes, co-design with disability communities, culturally situated evaluation, and comparative analysis of centralized and decentralized deployment models. Researchers should also examine how generative models can support rather than dilute local musical knowledge. The goal is not to reject AI in music education, but to build systems whose openness, accountability, and adaptability allow diverse learners to participate in musical creation on their own terms.

9. Conclusion

Human-AI co-creation has the potential to expand access to music education and enable inclusive creative experiences, but this potential is not an automatic consequence of technical progress. It depends on how systems are architected, governed, deployed, and evaluated. Accessible music education platforms must negotiate trade-offs between cloud scale and local control, between automation and creative agency, between personalization and privacy, and between standardized evaluation and culturally situated musical values. These trade-offs cannot be resolved by any single algorithm or interface feature. They require careful attention to modular architecture, participatory design, data governance, institutional capacity, and long-term policy.

A systems-level perspective reframes the central question. Instead of asking how AI can teach music more efficiently, the field should ask how AI can be embedded in sociotechnical arrangements that respect learner autonomy, educator expertise, cultural diversity, and disability justice. This shift implies new responsibilities for system designers, who must document assumptions and limitations; for institutions, which must build maintenance and training capacity; and for policymakers, who must create procurement and privacy frameworks that protect vulnerable learners. When these structural conditions are met, human-AI co-creation can become a meaningful part of an accessible and inclusive musical life, not a substitute for the human relationships that make music education transformative.

References

1. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-13). Association for Computing Machinery.

2. Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-16). Association for Computing Machinery.

3. Miranda, E. R., & Wanderley, M. M. (2006). New digital musical instruments: Control and interaction beyond the keyboard. A-R Editions.

4. Fiebrink, R., & Cook, P. R. (2010). The Wekinator: A system for real-time, interactive machine learning in music. In Proceedings of the 2010 International Computer Music Conference (pp. 366-369). International Computer Music Association.

5. Holland, S. (2000). Artificial intelligence in music education: A critical review. In E. R. Miranda (Ed.), Readings in music and artificial intelligence (pp. 239-274). Harwood Academic.

6. Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. Pearson Education.

7. Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.

8. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689-707.

9. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research, Vol. 81, pp. 77-91). PMLR.

10. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 59-68). Association for Computing Machinery.

11. Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

12. Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

13. Williamson, B. (2017). Big data in education: The digital future of learning, policy and practice. SAGE Publications.

14. Wang, Z., Ma, L., Jin, Y., Feng, Y., Pan, X., Ji, S., & Zhang, K. (2025, August). AI-assisted human-pet artistic musical co-creation for wellness therapy. In Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (pp. 10216-10224).

15. Benetos, E., Dixon, S., Duan, Z., & Ewert, S. (2019). Automatic music transcription: An overview. IEEE Signal Processing Magazine, 36(1), 20-30.

16. Sturm, B. L., Ben-Tal, O., Monaghan, Ú., Collins, N., Herremans, D., Chew, E., Hadjeres, G., Déguernel, K., & Pachet, F. (2019). Machine learning research that matters for music creation: A case study. Journal of New Music Research, 48(1), 36-55.

17. Keller, D., & Lazzarini, V. (2017). Ecologically grounded creative practice in ubiquitous music. Organised Sound, 22(1), 61-73.

18. Susskind, R., & Susskind, D. (2015). The future of the professions: How technology will transform the work of human experts. Oxford University Press.

19. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610-623). Association for Computing Machinery.

20. Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge University Press.