Who Gets Heard? Rethinking Fairness in AI for Music Systems

Atharva Mehta[1*] **Shivam Chauhan**[1*] **Megha Sharma**[1] **Gus Xia**[1]
Kaustuv Kanti Ganguli[2] **Nishanth Chandran**[3] **Zeerak Talat**[4] **Monojit Choudhury**[1]
[1]MBZUAI, UAE [2]Zayed University, UAE [3]Microsoft, India [4]University of Edinburgh, UK

{atharva.mehta, shivam.chauhan, megha.sharma}@mbzuai.ac.ae
{gus.xia, monojit.choudhury}@mbzuai.ac.ae
kaustuvkanti.ganguli@zu.ac.ae nichandr@microsoft.com z@zeerak.org

Abstract

In recent years, the music research community has examined risks of AI models for music, with generative AI models in particular, raised concerns about copyright, deepfakes, and transparency. In our work, we raise concerns about cultural and genre biases in AI for music systems (music-AI systems) which affect stakeholders—including *creators*, *distributors*, and *listeners*—shaping representation in AI for music. These biases can misrepresent marginalized traditions, especially from the Global South, producing inauthentic outputs (e.g., distorted ragas) that reduces *creators'* trust on these systems. Such harms risk reinforcing biases, limiting creativity, and contributing to cultural erasure. To address this, we offer recommendations at dataset, model and interface level in music-AI systems.

1 Introduction

AI models are rapidly transforming the music landscape with generative AI models that generate indistinguishable music compared to human-composed music[27]. State-of-the-art generative music models have improved significantly in automatic evaluation metrics and human evaluations [13, 31]; further bringing the generated music samples closer to human-composed music. The increasing capability of generative systems to produce music across diverse genres and perform strongly in human evaluations has motivated the development of advanced composition tools that integrate generative AI models. These advances enable new forms of expression, providing a new interface that can lower access barriers, e.g., allowing a form of "vibe music composition" that allows people to compose music without technical skills with instruments or in-depth knowledge of music [19]. Due to the ease of composition, researchers have recently raised concerns about the use of generative AI models for music generation (generative music models) [e.g., 5, 34, 14, 28, 36, 32], highlighting risks such as copyright infringement, undetected deep-fakes, market saturation, legal compliance, and nonadherence to human requests and preferences. Rising concerns among researchers and artists have led to concerted efforts to strengthen the safety of generative music models, addressing issues such as AI-generated music detection[11, 33], deepfake detection[1], and audio watermarking[15, 35, 26]. However, from the perspective of fairness (as defined previously in NLP [10]), very limited research has examined representational bias and cultural inclusion in music datasets and generative models. Similar concerns arose earlier in NLP: Bender et al. [6] showed that larger datasets do not ensure diversity, often overrepresenting harmful views such as misogyny and white supremacy, while Joshi et al. [23] highlighted disparities in language inclusion and resources. These findings led to the creation of bias detection benchmarks [16] and responsible AI frameworks [4] to mitigate stereotypes

[*]Equal Contribution

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI for Music.

and systemic inequities. Similarly, for music, Mehta et al. [29, 30] shows a huge skew in global representation[2] with a decreasing focus on Global South cultures, including South Asia, South East Asia, Oceania, and the Middle East, as well as genres like country & folk music, based on different definitions for fairness. Their analysis draws attention to the potential implications of such systemic biases, including widening economic disparities, limited global creativity in music, and, eventually, cultural erosion on a global scale. To fully address these implications, it is crucial to recognize that such biases are not accidental or harmless; they arise from long-standing patterns of cultural dominance, market-focused choices, and platforms that favor certain types of music while leaving others out [39]. Adding to this complexity is the long-tailed and evolving nature of music itself. Musical genres are fluid, hybrid, and constantly emerging in local, diaspora, and transnational contexts [20, 12]. It is practically and epistemologically impossible for any AI system to equally represent and capture the entirety of global musical expressions. Although these studies are crucial in understanding the metalevel effects of biases in music-AI systems, they do not capture the direct consequences of such biases on the diverse stakeholders in the music ecosystem, including *creators, distributors, platforms*, and *listeners*. With the integration of AI into composition and distribution pipelines, the roles and performance of each stakeholder are being reconfigured.

Thus, we examine the impact of biased representation of music in AI systems on different stakeholders in the music-AI ecosystem and discuss potential interventions at the data and interface levels to address misrepresentation and underrepresentation of musical traditions. Building on this, our goal is to guide future research at the intersection of generative AI, creativity, and social responsibility by addressing the following questions: (RQ1) Who are the stakeholders in the music-AI ecosystem and what are implications of representational biases on the stakeholders? (RQ2) What design choices and technical challenges in music-AI systems shape how representational biases emerge and affect stakeholders? (RQ3) Given the long tail of genres and the evolving nature of music, what system-, dataset- and model-level strategies can ensure fairness when equal representation is impractical?

We begin by identifying key stakeholders in the music-AI ecosystem, including *composers, listeners, distributors, teachers*, and *students*. We conducted unstructured interviews [42] with researchers, practitioners, and professionals in music and AI to understand risks associated with representational bias and fairness, particularly around cultural inclusion, misrepresentation, and their implications for stakeholders. Based on these discussions, we analyze each implication's impact on every stakeholder. In addition, to mitigate biases in music-AI systems, we propose a set of recommendations and open research questions for music & AI community. We situate music-AI within broader discussions of AI ethics and fairness, arguing that development must move beyond efficiency to support diverse and equitable creative expression.

2 RQ1 (i): Who are the stakeholders?

Stakeholders in music have long shaped the way music is produced, distributed, and experienced. Although their core roles remain, *composers*, *listener*s, *distributors*, *educators*, and *students*, their practices and interactions continue to evolve with technological change. Even though a single person in the music and AI ecosystem has the flexibility of performing multiple roles, we focus solely on the core roles themselves assuming that each person performs a single role. **Creators (composers and performers)** *Creators (composers and performers)* are the ones who write a musical piece in any form. At the individual level, this includes bands, DJs, and singer-songwriters; at the group level, cultural communities and traditions that preserve music as identity and heritage. To assist composers in compositions, AI is integrated into Digital Audio Workstations (DAWs), enabling creative ideas, effectively editing music with minimal effort, and enhancing live performances by providing accompaniment to artists in various forms, overall reducing the cost of music production. **Distributors and marketing professionals** *Distributors* and marketing professionals bridge *creators* and *listeners* (discussed in detail later). They can facilitate audience targeting for *creators* and enhance *listener* experience through preference-based distribution and marketing. Individually, they may be freelance promoters, curators, or bloggers; collectively, they include record labels, streaming platforms (Spotify), and large promotion networks (SoundOn by TikTok). Distributors and marketing professionals make use of recommendation systems based on Non-generative AI algorithms enabling *listener*-personalized recommendations, and market analysis, while generative models can create

[2]Throughout the remainder of the paper, we use "bias" and "representational bias with respect to data and system output representation interchangeably.,

promotional materials such as short clips or videos tailored to the target market. **Listeners** *Listeners* are those who listen to music, often with varying degrees of attention and engagement. *Listeners* experience and evaluate music both individually (e.g., through apps, study playlists, or background use) and collectively (fan communities, concert audiences). Non-generative AI recommendation systems aim to personalize and improve music discovery such as Spotify's Discover Weekly, helping *listeners* with playlist generation and context-aware recommendations. **Teachers and experts** *Teachers and experts* (music educators, scholars, and practitioners) play a vital role in preserving and transmitting musical knowledge. Individually, they mentor *students* and guide practice; at the group level, institutions and musicological communities curate archives, curricula, and research such as The Royal Academy of Music. AI tools can improve pedagogy and analysis, facilitate genre and music structure discovery and evolution, and provide new insights into human creativity and artistic expression. **Students** *Students* are learners who engage with music through formal or informal education, contributing to the preservation and evolution of cultural and creative traditions. At the individual level, they may study instruments, practice specific genres, or experiment with new forms. At the collective level, they form ensembles, learning communities, or cultural cohorts that sustain musical knowledge across generations. AI systems can support *students* by providing feedback on performance (eg, Tonara), generating accompaniment, offering interactive learning tools, and simulating performance environments.

3 RQ1 (ii): Implications of representation bias on stakeholders

Now that we understand the different stakeholders in the music-AI ecosystem, we can examine the social, cultural, and ethical issues that arise when these groups are overlooked [38, 21]. **Misrepresentation Harms** Misrepresentation of culturally specific or underrepresented music can distort productions: *creators* face a mischaracterization that undermines authenticity[3] and trust further undermining artist performances; *listeners* intend to explore diversity, but the available content misrepresents it, shaping their perception incorrectly; and *teachers* and *students* can unknowingly propagate these distortions in education if they use such tools for creating educational content[e.g., 7, 24, 41, 22, 18]. **Homogenization and Exposure Bias** Homogenization and exposure bias compound the problem. Probabilistic models produce uniform outputs shaped by dominant datasets, and recommendation systems reinforce *listener* exposure to familiar music [40, 25]. *Creators* struggle to innovate or express cultural depth; small-scale *distributors* and labels who produce and distribute underrepresented music are marginalized; *listeners* develop narrow preferences, limiting discovery; and *students*, who need exposure to diverse musical genres, limit their creative discovery process. **Cultural Erosion & Widening Economic Disparities** Generative models can erode cultural identity by homogenizing traditional music, reducing access to distinctive regional forms [37, 9, 3]. Due to lack of music discovery often the under-represented genres and practitioners of such genres face limited attention from *listeners*; *distributors* may prioritize popular genres over traditional ones, further drawing these genres to a position of disadvantage; eventually widening the economic disparity between *creators* composing popular and under-represented genres. This leads to a snowball effect wherein lesser *creators* from the community pursue these genres as career due to lack of financial resources, eventually which can lead to erosion of such genres. *Listeners* encounter barriers to engaging with authentic musical cultures, leaving scarce options for music discovery, and *students* risk learning culturally homogenized material. **Opaque Training Processes and Datasets** Closed source models like Udio and many open source models including MusicLM[2], etc are trained on closed datasets often not accessible to the community. This leads to training models on uncurated datasets with unclear composition and intent [17]. Opaque datasets and models make it impossible to trace where music originates. As a result, AI-generated music can mimic styles or compositions without proper attribution, and human *creators* lose recognition for their work. As a result, *listeners*, *teachers*, and *students* may unknowingly treat AI-generated content as authentic, further eroding appreciation of human creativity.

4 RQ2: Design choices & Technical Challenges

The ethical implications discussed above, such as cultural bias and misrepresentation, often stem from technical gaps embedded in generative systems that stem from how data, representations, and

[3]Expression of the society's historically important musical traditions, repertoires, and practices.

interfaces are designed. Although these appear system-centered, they ultimately shape the socio-cultural gaps as discussed in Section 3, shaping the kind of music people can generate, access, and experience. **First**, genre labels in datasets and interfaces are often flattened into ambiguous categories such as world or ethnic, collapsing rich traditions into catch-alls and reinforcing cultural erasure. This reflects a Western preoccupation with classification, where hyperspecific microgenres (e.g., melodic house) coexist alongside the homogenization of non-Western forms, which are treated as undifferentiated. **Second**, training data rarely retains regional or cultural metadata (e.g., place of origin, instruments, performance context, or linguistic association), making it impossible for models to distinguish between culturally rooted genres and generic patterns. **Third**, symbolic formats like MIDI, while compact and portable, encode Western tonal and rhythmic defaults, marginalizing microtonality, heterophony, and non-metric rhythms that fall outside its representational frame. **Fourth**, prompt interfaces often guide users through culturally narrow defaults, such as mood and genre tags such as cinematic or relaxing, which, combined with overreliance on English metadata and tags, further restrict exploratory access to underrepresented traditions. Together, these design choices produce a structurally homogenized music space, where non-Western and fluid traditions are not only underrepresented but actively constrained by the infrastructures of generation.

5 RQ3: Recommendations

To address technical shortcomings in trained music-AI systems, we propose recommendations across three levels—datasets, models, and system/interfaces—aimed at improving transparency and bridging the technical gaps that contribute to socio-cultural inequities in music-AI. **Dataset level**. Music datasets should be better documented following practices like datasheets and data statements [8, 6] including origin, recording conditions, *artists/creators* involved (to account for attribution), instruments, genres, regions, style of play and permissions. Data sets should be audited for all genres to estimate the representation of different genres similar to Mehta et al. [29] or using entropy to measure genre diversity and supported with richer labels rather than vague tags like "world music". Where sensitive or sacred recordings are involved, communities should be consulted, consent obtained, and material labeled to prevent casual misuse. **Model level**. Models should support traceability, letting users see what data influenced an output, which aids transparency, credit, and user trust. Training methods must reduce overfitting to dominant genres (e.g., reweighting, balanced sampling) and include parameters to explore underrepresented styles. Evaluation should move beyond signal quality to account for genre diversity, style preservation, and cultural variation, even if such measures are still complex and emerging. **System/interface level**. Interfaces shape user access: genre and region controls should be simple, visible, and not hidden under “advanced settings.” Systems should avoid flattening labels like world music, instead surfacing specific terms (e.g., Gnawa, Taiko) and presenting contextual information about style, region, or instruments. Tool tips or links can guide users toward learning about traditions they generate, fostering awareness and inclusion. **Governance level.** Fairness requires structural measures: involving musicians and cultural *teachers and experts* in system design, defining guidelines for use of sensitive music, enabling opt-out mechanisms for *creators*, and supporting compensation or licensing models where appropriate. Policymakers should extend the regulation of AI to include cultural expression and creative labor.

6 Conclusion & Future Work

In this paper, we identified key stakeholders in the music-AI ecosystem and examined the implications of representational bias for each group. We highlighted socio-cultural and technical gaps that further enforce under-representation in music-AI systems and proposed recommendations across datasets, models, and interfaces. Looking ahead, several directions of future research open up: evaluating fairness and cultural representation in generated music; ensuring credit and consent for *creators (composers)* and communities with appropriate policies and governance; supporting cultural fidelity without reinforcing stereotypes; extending symbolic infrastructures like MIDI to non-Western traditions; and addressing language-driven biases in prompts and metadata. Together, these directions chart a path toward music-AI systems that are not only technically capable but also more inclusive and culturally respectful.

References

[1] Darius Afchar, Gabriel Meseguer-Brocal, and Romain Hennequin. Detecting music deepfakes is easy but actually hard. *arXiv preprint arXiv:2405.04181*, 2024.

[2] Andrea Agostinelli, Timo I Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, et al. Musiclm: Generating music from text. *arXiv preprint arXiv:2301.11325*, 2023.

[3] Iran Davar Ardalan, Amir Banifatemi, Fernando Gonzalez, Myles Ingram, Reza Moradinezhad, and Lucretia Williams. *AI for Community: Preserving Culture and Tradition*. CRC Press, 2025.

[4] Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, et al. Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*, 2022.

[5] Julia Barnett. The ethical implications of generative audio models: A systematic literature review. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, page 146–161, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400702310. doi: 10.1145/3600211.3604686. URL `https://doi.org/10.1145/3600211.3604686`.

[6] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623, 2021.

[7] Su Lin Blodgett and Brendan O'Connor. Racial disparity in natural language processing: A case study of social media african-american english. *arXiv preprint arXiv:1707.00061*, 2017.

[8] Karen L Boyd. Datasheets for datasets help ml engineers notice and understand ethical issues in training data. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–27, 2021.

[9] Philip Burnett, Erin Johnson-Williams, and Yvonne Liao. Music, empire, colonialism: sounding the archives, 2023.

[10] Monojit Choudhury and Amit Deshpande. How linguistically fair are multilingual pre-trained language models? *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(14): 12710–12718, May 2021. doi: 10.1609/aaai.v35i14.17505. URL `https://ojs.aaai.org/index.php/AAAI/article/view/17505`.

[11] Luca Comanducci, Paolo Bestagini, and Stefano Tubaro. Fakemusiccaps: A dataset for detection and attribution of synthetic music generated via text-to-music models. *Journal of Imaging*, 11(7), 2025. ISSN 2313-433X. doi: 10.3390/jimaging11070242. URL `https://www.mdpi.com/2313-433X/11/7/242`.

[12] John Connell and Chris Gibson. World music: deterritorializing place and identity. *Progress in Human Geography*, 28(3):342–361, 2004.

[13] Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez. Simple and controllable music generation. *Advances in Neural Information Processing Systems*, 36:47704–47720, 2023.

[14] Jacopo de Berardinis, Lorenzo Porcaro, Albert Merono-Penuela, Angelo Cangelosi, and Tess Buckley. Towards responsible ai music: an investigation of trustworthy features for creative systems. *arXiv preprint arXiv:2503.18814*, 2025.

[15] Pascal Epple, Igor Shilov, Bozhidar Stevanoski, and Yves-Alexandre de Montjoye. Watermarking training data of music generation models. *arXiv preprint arXiv:2412.08549*, 2024.

[16] Isabel O. Gallegos, Ryan A. Rossi, Joe Barrow, Md Mehrab Tanjim, Sungchul Kim, Franck Dernoncourt, Tong Yu, Ruiyi Zhang, and Nesreen K. Ahmed. Bias and fairness in large language models: A survey. *Computational Linguistics*, 50(3):1097–1179, 09 2024. ISSN 0891-2017. doi: 10.1162/coli_a_00524. URL `https://doi.org/10.1162/coli_a_00524`.

[17] Mary L Gray and Siddharth Suri. *Ghost work: How to stop Silicon Valley from building a new global underclass*. Harper Business, 2019.

[18] Kimia Hamidieh, Haoran Zhang, Walter Gerych, Thomas Hartvigsen, and Marzyeh Ghassemi. Identifying implicit social biases in vision-language models. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, pages 547–561, 2024.

[19] Dorien Herremans, Ching-Hua Chuan, and Elaine Chew. A functional taxonomy of music generation systems. *ACM Computing Surveys (CSUR)*, 50(5):1–30, 2017.

[20] Fabian Holt. *Genre in popular music*. University of Chicago Press, 2019.

[21] Sara Hooker. Moving beyond "algorithmic bias is a data problem". *Patterns*, 2(4), 2021.

[22] Jen-tse Huang, Jiantong Qin, Jianping Zhang, Youliang Yuan, Wenxuan Wang, and Jieyu Zhao. Visbias: Measuring explicit and implicit social biases in vision language models. *arXiv preprint arXiv:2503.07575*, 2025.

[23] Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. The state and fate of linguistic diversity and inclusion in the NLP world. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.560. URL `https://aclanthology.org/2020.acl-main.560/`.

[24] Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R Rickford, Dan Jurafsky, and Sharad Goel. Racial disparities in automated speech recognition. *Proceedings of the national academy of sciences*, 117(14):7684–7689, 2020.

[25] Vijay Kumar, Harit Pandya, and CV Jawahar. Identifying ragas in indian music. In *2014 22nd International Conference on Pattern Recognition*, pages 767–772. IEEE, 2014.

[26] Yixin Liu, Lie Lu, Jihui Jin, Lichao Sun, and Andrea Fanelli. Xattnmark: Learning robust audio watermarking with cross-attention. *arXiv preprint arXiv:2502.04230*, 2025.

[27] Iman Malik and Carl Henrik Ek. Neural translation of musical style. *arXiv preprint arXiv:1708.03535*, 2017.

[28] Jon McCormack, Toby Gifford, and Patrick Hutchings. Autonomy, authenticity, authorship and intention in computer generated art. In *International conference on computational intelligence in music, sound, art and design (part of EvoStar)*, pages 35–50. Springer, 2019.

[29] Atharva Mehta, Shivam Chauhan, and Monojit Choudhury. Missing melodies: Ai music generation and its" nearly" complete omission of the global south. *arXiv preprint arXiv:2412.04100*, 2024.

[30] Atharva Mehta, Shivam Chauhan, Amirbek Djanibekov, Atharva Kulkarni, Gus Xia, and Monojit Choudhury. Music for all: Representational bias and cross-cultural adaptability of music generation models. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 4569–4585, 2025.

[31] Jan Melechovsky, Zixun Guo, Deepanway Ghosal, Navonil Majumder, Dorien Herremans, and Soujanya Poria. Mustango: Toward controllable text-to-music generation. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 8286–8309, 2024.

[32] Benjamin A Morgan. Computing taste: algorithms and the makers of music recommendation. *International Journal of Music Business Research*, 12(2):79–81, 2023.

[33] Md Awsafur Rahman, Zaber Ibn Abdul Hakim, Najibul Haque Sarker, Bishmoy Paul, and Shaikh Anowarul Fattah. Sonics: Synthetic or not–identifying counterfeit songs. *arXiv preprint arXiv:2408.14080*, 2024.

[34] Matthew Sag. Copyright safety for generative ai. *Hous. L. Rev.*, 61:295, 2023.

[35] Robin San Roman, Pierre Fernandez, Antoine Deleforge, Yossi Adi, and Romain Serizel. Latent watermarking of audio generative models. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2025.

[36] Nick Seaver. Computing taste: Algorithms and the makers of music recommendation. In *Computing Taste*. University of Chicago Press, 2022.

[37] Linda Tuhiwai Smith. *Decolonizing methodologies: Research and indigenous peoples*. Bloomsbury Publishing, 2021.

[38] Nathalie A Smuha. Beyond the individual: governing ai's societal harm. *Internet Policy Review*, 10(3), 2021.

[39] Qinyuan Wang, Bruce Gu, He Zhang, and Yunfeng Li. Fairness of large music models: From a culturally diverse perspective. In *2024 IEEE 9th International Conference on Data Science in Cyberspace (DSC)*, pages 706–712. IEEE, 2024.

[40] F Yamada, Hiroshi Bando, et al. Traditional japanese gagaku music for contribution of human wellness. *Global Journal of Arts and Social Sciences*, 4(1):164, 2022.

[41] Xinzhi Zhang, Mary Frances Cotch, Asel Ryskulova, Susan A Primo, Parvathy Nair, Chiu-Fang Chou, Linda S Geiss, Lawrence E Barker, Amanda F Elliott, John E Crews, et al. Vision health disparities in the united states by race/ethnicity, education, and economic status: findings from two nationally representative surveys. *American journal of ophthalmology*, 154(6):S53–S62, 2012.

[42] Yan Zhang and Barbara M Wildemuth. Unstructured interviews. *Applications of social research methods to questions in information and library science*, 2:222–231, 2009.