

Missing Melodies: AI Music Generation and its "Nearly" Complete Omission of the *Global South*

Atharva Mehta^{*1}, Shivam Chauhan^{*1}, Monojit Choudhury¹

¹Mohamed bin Zayed University of Artificial Intelligence
{atharva.mehta, shivam.chauhan, monojit.choudhury}@mbzuai.ac.ae

Abstract

Recent advances in generative AI have sparked renewed interest and expanded possibilities for music generation. However, the performance and versatility of these systems across musical genres are heavily influenced by the availability of training data. We conducted an extensive analysis of over one million hours of audio datasets used in AI music generation research, and manually reviewed more than 200 papers from eleven prominent AI and music conferences and organizations to identify a critical gap in the fair representation and inclusion of the musical genres of the *Global South* in AI research.

Our findings reveal a stark imbalance: approximately 86% of the total dataset hours and over 88% of researchers focus primarily on music from the *Global North*. However, around 40% of these datasets include some form of non-Western music, and genres from the *Global South* account for only 14.6% of the data. Furthermore, approximately 51% of the papers surveyed concentrate on symbolic music generation, a method that often fails to capture the cultural nuances inherent in music from regions such as South Asia, the Middle East, and Africa.

As AI increasingly shapes the creation and dissemination of music, the significant underrepresentation of music genres in datasets and research presents a serious threat to global musical diversity, including evaluating music generation models, economic disparities, and reinforcing biases in generative music models. We also propose some important steps to mitigate these risks and foster a more inclusive future for AI-driven music generation.

Introduction

Music has always been a crucial element in representing the traditions of different communities worldwide (Savage 2019). With recent developments in AI, particularly through the use of deep learning models (Copet et al. 2024; Schneider et al. 2024; Melechovsky et al. 2024; Agostinelli et al. 2023), there has been a dramatic improvement in generating music automatically. These advancements have led to the creation of various AI-driven music platforms, such as

Jukebox (Dhariwal et al. 2020), Suno¹, Udio², Boomy³, and Soundraw⁴, giving users the ability to create music based on their preferences using text prompts. In terms of generation quality and performance, AI systems have significantly improved on automatic evaluation metrics as well as human evaluations, enabling more human-like and controlled music generation, indicating a significant improvement in their capabilities (Melechovsky et al. 2024; Agostinelli et al. 2023; Schneider et al. 2024; Copet et al. 2024). Due to the ease of use, people can now "vibe compose" musical pieces with minimal musical knowledge. Dysart (2018) noted, "AI-generated music could easily surpass the amount of music that has ever been recorded, due to how quickly it can be produced." In terms of its consumption, Mubert, an online AI music composition website, reported that by mid-2023 its users had streamed more than 100 million AI-generated tracks⁵. With growing compute and efficient models, systems like Mousai can generate 43-second music clips in 49.2 seconds using just 5.04 GB GPU memory, enabling round-the-clock production far beyond human efficiency.

As noted in past studies (Tao et al. 2024; Ungruh et al. 2024), despite impressive technological progress, biases remain prevalent in AI music generation systems. These include popularity, training data bias, and limited research interest, which leads to under-representation of musical genres from the *Global South* regions of the world. For instance, the model described in Melechovsky et al. (2024); Copet et al. (2024); Schneider et al. (2024) often defaults to Western tonal and rhythmic structures when tasked with generating non-Western music such as Indian classical or traditional Middle Eastern genres, let alone lesser-known styles and genres such as the Baul⁶ and instruments such as the Gonje⁷.

¹<https://suno.com/>

²<https://www.udio.com/>

³<https://boomy.com/>

⁴<https://soundraw.io/>

⁵<https://mubert.com/blog/an-in-depth-study-into-ai-music-its-creators-composers-and-adopters>

⁶The Bauls are mystic minstrels from the Bengal region spread across India and Bangladesh, blending Sufism and Vaishnavism in songs about the love between the human soul and a personal god within.

⁷The Gonje is a one-stringed West African fiddle, often made with a snakeskin-covered gourd and horsehair string, played solo

^{*}These authors contributed equally.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

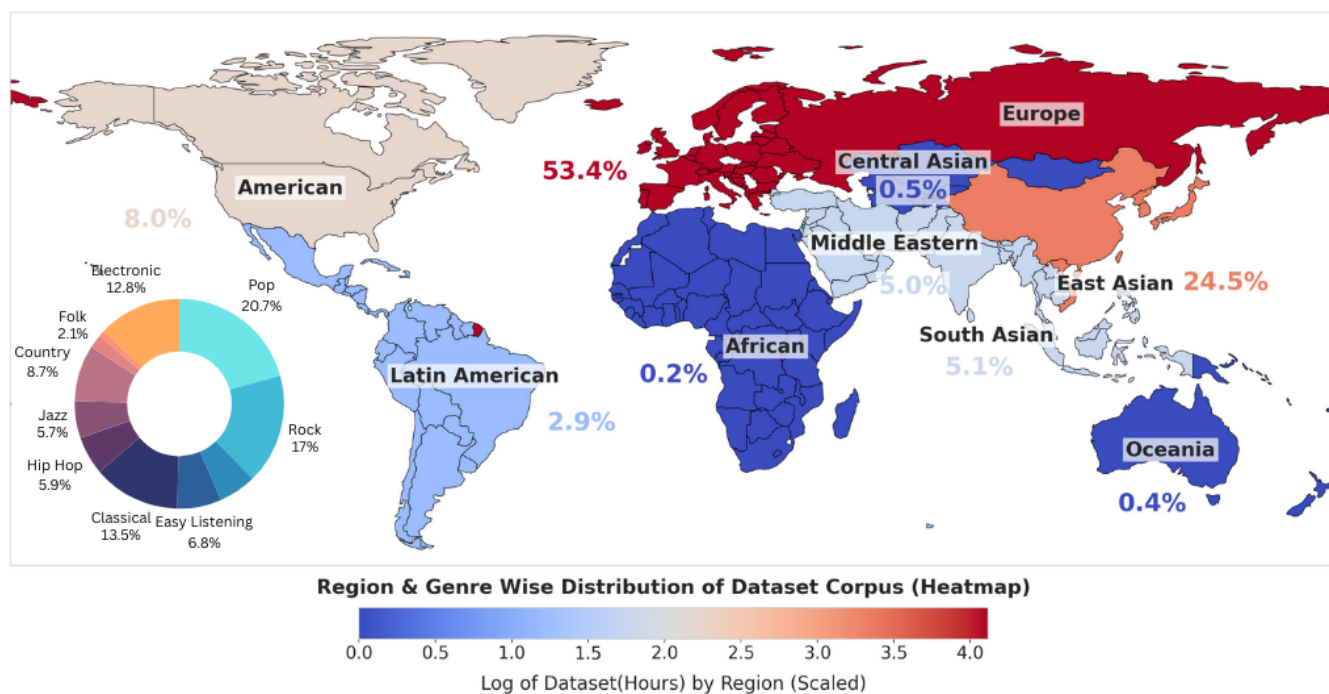


Figure 1: Global Divide in AI Music Datasets: The heatmap shows the stark imbalance in regional musical styles (on the map) and genre (pie chart) representation in popular music datasets.

As a result, a generated piece intended to mimic an Indian raga may sound like a Western pop melody played on a sitar. Similarly, SunoAI, when attempting to generate Maqamat⁸ music of the Middle East, may round off the microtones to the nearest Western equivalent, resulting in a piece that lacks the distinctive sound of Arabic music.

As generative models continue to gain traction in the field of music generation, the misrepresentation and under-representation of the musical genres of the “global majority” poses a significant threat to the inclusion of musical genres from around the world. The skewed distribution in datasets, reflected in model outputs, can lead to several issues, including cultural homogenization, reinforcement of Western culture dominance⁹, misrepresentation of musical styles, and most importantly, gradual decline leading to the endangerment and disappearance of many musical genres (Tan 2021; Lund 2019; Team 2023). Drawing inspiration from studies such as Joshi et al. (2020) and Bender and Friedman (2018), which systematically analyze the under-representation of languages of the global majority in the field of natural language processing, we attempt to analyze the under-representation of musical genres in the field of AI music generation.

or in ensembles with other traditional instruments.

⁸Maqam, in traditional Arabic music, is a melodic mode system defining pitches, patterns, and improvisation, central to Arabian art music, with 72 heptatonic scales.

⁹<https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>

In this paper, we examine the distribution of musical genres and regional music from around the world in the datasets that are used to train models and conduct research in AI-music. We also study and contrast the distribution of participation of researchers from different parts of the world to the latter. This allows us to examine whether regions under-represented in datasets are also underrepresented in authorship, further reinforcing existing imbalances. To this end, we analyze many popular datasets and papers from eleven AI conferences and organizations and contrast these findings with the digital (online) consumption & availability and ideal (offline) proxies of population. Our study reveals almost a complete omission of the musical styles of the *Global South* in the AI world (Figure 1 and Table 1). Approximately 86% of the total hours in available datasets and 93% of papers focus majorly on the music of the *Global North*. In contrast, only 14.6% of the total hours of music data and 12.1% of papers are first-authored by researchers from South Asian, Middle Eastern, Oceania, Central Asian, Latin American, and African countries. This distributional skew in the training data is reflected in the quality of generated samples from several popular systems, including Jukebox, Suno, and Udio (examples are available for listening on our GitHub page¹⁰). We discuss several implications of this representational skew on the resulting quality of the generated music, research, and society, including concerns ranging from improper evaluations of generative models and limited creativity and diversity in generative AI-assisted music composi-

¹⁰<https://atharva20038.github.io/aimusicexamples.github.io/>

tion for the underrepresented genres to cultural erosion.

The article is organized as follows: The next section, Framework and Research Questions, formalizes the key concepts and describes our data collection and analysis pipelines. We then present our findings in Results, discuss their broader implications and possible mitigation strategies in Discussion, and conclude in Conclusion.

Research Methods and Materials

Definitions & Research Questions

We begin by mathematically defining fairness and representation.

Representation. Suppose $G = \{g_1, g_2, \dots\}$ is the set of genres (e.g., *classical*, *folk*) and $R = \{r_1, r_2, \dots\}$ is the set of regions (e.g., *Oceania*, *Latin America*). We define a musical style as $s = \langle g, r \rangle \in G \times R$, that is, a combination of a genre and a region. The dataset-based and publication-based representation measures are

$$\delta_g = \frac{D_g}{\sum_{g' \in G} D_{g'}}, \quad \delta_r = \frac{D_r}{\sum_{r' \in R} D_{r'}},$$

$$\rho_s = \frac{P_s}{\sum_{s' \in G \times R} P_{s'}}, \quad \alpha_r = \frac{P_r}{\sum_{r' \in R} P_{r'}}.$$

Here, D_g and D_r represent the dataset sizes in hours for a specific genre g or region r , respectively; P_s is the number of surveyed papers focusing on style s ; P_r is the number of surveyed papers with a first author from region r .

The *performance* of a model in a musical style is proportional to the amount of music data from that style that was used to train the model. This seemingly self-evident fact has also been empirically confirmed in numerous studies in machine learning literature. See Henning et al. (2023), for example, for a survey of such studies for language data and resultant performance disparity in LLMs. It follows that fair or equitable performance of models across musical styles implies an equivalently fair or equitable representation of the style (i.e., D_g or D_r) in the dataset. However, there is no single or universal notion of *fairness*. For our purposes, we adopt the Rawlsian notion of fairness (Rawls 1958) defined as the principle of least difference. More precisely,

Fairness. A music generation system is said to be fair across styles if disparities in representation across genres and regions are small. Using the dataset-based representation measures, we quantify such disparities by

$$\max_{g \in G} \delta_g - \min_{g \in G} \delta_g$$

and

$$\max_{r \in R} \delta_r - \min_{r \in R} \delta_r,$$

and regard smaller values as indicating greater fairness.

Similarly, fairness of representation can be defined with respect to the publication-based representation measures ρ_s and α_r .

Fairness may also be defined from a utilitarian perspective, which argues for representation of styles in datasets and publications in proportion to their real-world utility. Although the utility of a musical style is difficult to estimate

directly, it may be approximated through real-world production and consumption patterns. To this end, we propose two *proxies for the real-world utility* of musical styles: (i) availability and consumption patterns in the digital space, and (ii) the population of the corresponding region.

Digital and Ideal Proxies of Utility We define digital proxies, π_{dl} , as a collection of digital music listenership and π_{da} , as the digital availability of styles and regions. Similarly, we define the ideal proxy, π_{id} , as the regional population: $\langle \pi_{pp} \rangle$.

A fair representation in a utilitarian sense would imply a high positive correlation between the digital and ideal proxies of a style with its data-based and publication-based representations. We call population the *ideal* proxy because the digital space is also an unfair or skewed representation of the real world, and ideally, the number of people interested in consuming and producing a particular style of music should be considered when estimating the true utility of a style. We must also recognize the limitations of *population* as a proxy for real-world consumption or popularity of a musical style, especially because several genres and traditions of music coexist among all populations, and their relative prevalence must be factored in. Nevertheless, due to the lack of better proxies, the population has been commonly used as an important indicator for studying disparities in the digital and technological world (Florida, Mellander, and Stolarick 2010; Joshi et al. 2020). Building on these definitions, we formulate the following research questions:

1. What is the extent and fairness of the representation of various genres (δ_g) and regional music (δ_r) in the datasets?
2. What is the extent and fairness of the representation of styles and regions in the research as captured by (α_r) and (ρ_s)?
3. What is the correlation between the real-world proxies (π_{dl} , π_{da} and π_{id}) and the data and publication-based representations?
4. What are the implications of skewed representations in AI-music research on the global music ecosystem?

Data Collection

To get our initial pool of papers, we employed a multi-stage, keyword-based paper selection method, leveraging the Scholarly package (Cholewiak et al. 2021) to gather approximately 5000 papers. This included up to 1000 papers per query, using broad search terms such as “music,” “music generation,” “non-western music,” “MIDI,” and “symbolic music.” We then refined our selection by focusing on papers presented at eleven major conferences, including AAAI, ACM, EUSIPCO, EURASIP, ICASSP, ICML, IJCAI, ISMIR, NeurIPS, NIME, and SMC, chosen based on their popularity and prestige, narrowing our pool to around 800 papers.

Dataset Papers To identify papers proposing datasets, we read through the title and abstract of each paper. This led to a set of 152 papers¹¹ proposing new datasets with a total of over one million hours of music. These datasets were

¹¹<https://github.com/atharva20038/aimusicexamples.github.io/blob/master/Surveyed/%20Papers/Dataset-Papers.md>

manually annotated for the region and genres covered, total hours of music data, and whether the dataset was annotated with other details (such as instruments, genre, and style). Papers that directly provided details of the distribution of data points across genres and regions were analyzed with the already available statistics. Unfortunately, several datasets did not offer substantial details necessary for our study. If a dataset contained over 10,000 hours of audio, we analyzed the metadata for each file to extract genre and regional tags which were then mapped to one of the following genre categories (G) – Folk, Electronic, Pop, Classical, Rock, Easy Listening, Hip-Hop, Jazz, Country, Reggae & Soul, Experimental and Others. The region tags were mapped to our predefined region categories (R) – Europe, South Asia, East Asia, America, Africa, Latin America, Oceania, Middle East, and Central Asia (see Figure 1). However, when the genre and region were not explicitly mentioned in either the paper or the metadata, we did not make any assumptions; thus, 7.9% of the datasets totaling 5,772 hours were excluded from our analysis. The remaining annotated data in hours for each genre (D_g) and region (D_r) are used for further analysis.

Research Papers We conducted another keyword-based filtering on the titles and abstracts of the papers (with keywords such as "generative AI", "music generation", "auto-encoders", and "transformers") to further restrict the pool to those that propose generative AI techniques for music generation. This led us to a pool of 244 papers¹² that were further annotated for genres using the same method used to annotate datasets and further form our set of published papers with style annotations (P_s). Due to the lack of accurate data on region-wise information on genres such as hours of data on Kenyan folk music, which would often be mapped to "folk" in most papers, we simply take the summation over all regions to represent P_s ; i.e., $P_s = \sum_{r \in R} P_{s,r}$. For regional representation, we focused on the first author's institutional affiliation, based on the location of the institution rather than the geographical focus of the study in the paper (P_r). The institution's country for the first author was mapped to R . This analysis focuses on first authors, as they are often regarded as the lead contributors who take primary responsibility for the research and dedicate the most time and effort to its development. For instance, if the first author was affiliated with a university in France, the paper was labeled as European, ensuring a clearer picture of region-wise research representation.

Proxy Estimation For digital music, we draw on data from *SoundCharts*¹³ and the *MusicBrainz* API¹⁴. In *SoundCharts*, we examine the YouTube views of the top 50 most-viewed songs from each region, summing the total views to evaluate music listenership across regions. While the portal lacks data from *Australia*, and the *Middle East*, and does not provide a detailed breakdown within *Asia*, it offers a rough comparison between the music from *Global North*

and *South*. The *MusicBrainz* API, on the other hand, codes the region where a song was recorded, offering an estimate of the number of songs produced in each region that exist in the digital space. For the ideal space, we use population as a proxy and we collect its data from Wikipedia¹⁵.

Findings

In this section, we present the findings of our analysis of datasets and research papers.

RQ1: Representativeness of the Datasets

Genre-wise Distribution (δ_g): As shown in Figure 1, *Pop* music has the highest (20.7%) representation followed by *Rock* (17%) and *Classical* (13.5%) genres. *Country*, *hip-hop*, *blues* and *jazz* have a moderate (more than 5%) representation while *Folk* and *experimental* music have a low representation of 2.1%. The other genres have minimal to no representation ($\leq 1\%$), which includes music for Children, *Indie-music*, and region-specific genres.

Table 1 shows that *Pop* music forms 200K+ hours of the collection while *Folk* music constitutes only 20K hours of the collection. *Pop*, *Rock*, *Classical* and *Electronic* music genres each have more than 10% representation and more than 100K hours of data available.

Region-wise Distribution (δ_r): We find that more than 6k hours of music in the research belongs to *European* music and only 28 hours of music belong to *African* music, as shown in Table 1. *European*, *East Asian* and *American* music are well represented in the datasets, cumulatively having 85.9% representation. On the other hand, *South Asian* and *Middle Eastern* music has low representation, approximately 5% each. *Central Asian* and *African* music have less than 1% representation as depicted in Figure 1 showing they have extremely low representation. Notably, music research datasets often emphasize European classical music, which are often copyright-free, while US pop, country, and rock receive less attention, contributing to lower US participation in dataset creation compared to Europe. As can be seen in the figure, a striking picture emerges when we plot the heatmap; *Global South* is marked in blue, while *Global North* appears in red, highlighting the imbalance in representation of regions in music research datasets.

RQ2: Representativeness of Research

Genre-wise Distribution (ρ_s): Figure 2 shows that *Classical* music is the most prominent genre in music generation research, accounting for 36.5% representation in research. *Electronic* music, with a substantial 14.9%, and *Pop*, at 13.3%, follow as the next most studied genres. These genres likely benefit from a high volume of accessible data and established digital formats, making them more conducive to computational research. *Avant-garde & Experimental* genres have significant representation as well (10.2%), underscoring interest in exploring unconventional and innovative sounds within AI-driven music generation. Confer-

¹²<https://github.com/atharva20038/aimusicexamples.github.io/blob/master/Surveyed%20Papers/Music-Papers.md>

¹³<https://soundcharts.com/>

¹⁴<https://musicbrainz.org/>

¹⁵https://en.wikipedia.org/wiki/List_of_continents_and_continental_subregions_by_population

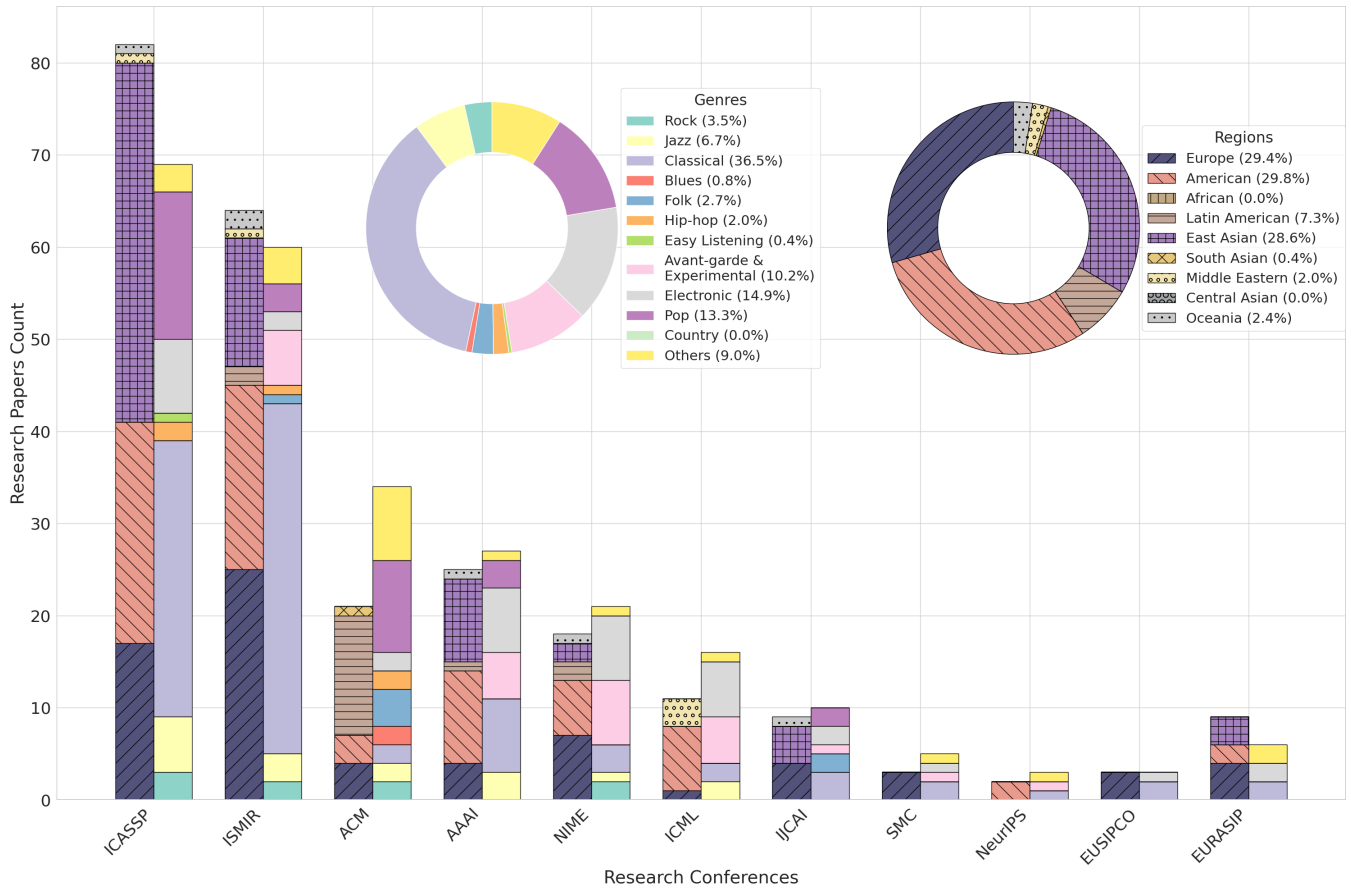


Figure 2: Genre and Regional Representation (α_r/ρ_s) in AI Music Research by Conferences: The bar plot illustrates the conference-wise representation of genre and region (derived respectively from the music datasets and the first author’s institutional affiliation). The pie charts show the overall representation by genre (left) and region (right).

ences such as *ICASSP* and *ISMIR* have a strong representation of *Classical*, *Pop*, and *Electronic* genres, while *NIME* and *AAAI* showcase a diverse genre distribution, including *Avant-garde & Experimental* genres, indicating these venues’ openness to exploring other genres also in music generation. Genres like *Jazz* (6.7%), *Rock* (3.5%), and *Folk* (2.7%) have moderate representation, suggesting a niche but valuable exploration of these genres in AI music research. *Country* and *Easy Listening* are among the most underrepresented genres in music generation research.

Region-wise Distribution (α_r): Figure 2 highlights significant disparities in research focus between genres of music and affiliations of researchers, underscoring the underrepresentation of the *Global South*. Cumulatively, representation of all regions in the *Global North* is approximately 88%, with Europe having the highest representation. In contrast, *Global South* has only about 12% representation. This severe under-representation of *Global South* researchers threatens the preservation of cultural heritage from regions such as *Africa*, *Middle East*, and *South Asia*, which collectively account for less than 2.4% representation.

Table 1 shows that *East Asia* (84 papers), *Europe* (70 papers), and *America* (75 papers) are leading in terms of

researcher-affiliations, which is likely due to the strong music industries in these regions. However, regions like *Africa*, *Latin America*, and *Oceania* show minimal or no representation. At ACM venues including *MMAAsia*¹⁶, *ICMR*¹⁷, *AIMLSystems*¹⁸, *UMAP*¹⁹, *C&C*²⁰, *TKDD*²¹, and *TOMM*²², there is a notable emphasis on *Pop*, with most researchers affiliated with institutions in *East Asia*. We found only one ACM paper whose first author was affiliated with an institution in *South Asia*.

RQ3: Correlation between Representation and Utility

Upon analyzing the digital and real-world proxies, the data reveal significant disparities across regions: *American* music leads digital consumption with 150 billion views, yet *Euro-*

¹⁶<https://www.acmmmasia.org/>

¹⁷<https://www.icmr2024.org/>

¹⁸<https://www.aimlsystems.org/>

¹⁹<https://www.um.org/umap2024/>

²⁰<https://cc.acm.org/2022/>

²¹<https://dl.acm.org/journal/TKDD>

²²<https://dl.acm.org/journal/tomm>

Genre	MB	SC	Papers	Dur.
Pop	49955	168.32	34	228.26
Rock	70823	106.89	9	186.67
Electronic	14025	115.73	38	140.43
Classical	8876	13.72	94	148.89
Country	2906	31.70	–	95.77
Hip-hop	22562	87.87	5	64.38
Jazz	5663	11.38	17	62.20
Blues	13979	78.12	2	64.06
Easy Listening	486	1.00	2	74.41
Folk	9390	52.75	7	22.82
Avant-garde/Exp.	2655	–	26	11.32
Others	9295	–	23	0.94

(a) Genre-wise distribution

Region	Pop	MB	SC	Papers	Dur.
European	0.75	14433	2.38	70	6127.92
East Asian	1.66	662	58.39	84	2817.73
South Asian	2.07	671	–	2	588.78
Central Asian	0.08	945	–	0	57.01
American	0.58	28823	149.24	75	921.84
Latin American	0.66	2265	82.54	5	332.86
Oceania	0.05	484	–	3	41.99
African	1.22	945	4.49	0	27.50
Middle Eastern	0.47	1366	–	5	569.86

(b) Region-wise distribution

Table 1: Distribution of datasets across (a) genres and (b) regions. Metrics: Duration (Dur., 10^3 hours in (a), hours in (b)), MusicBrainz (MB, songs), SoundCharts (SC, billions), papers, and population (billions, regions only).

pean music, with only 2.38 billion views on SoundCharts, dominates research datasets (6,127.92 hours). Meanwhile, *South Asia*, with the largest population of 2.07 billion, is severely underrepresented in research (588.78 hours) and digital catalogs (671 songs). Similarly, *Africa*, despite a population of 1.22 billion, has minimal representation (27.50 hours) and digital catalogs (945 songs). *East Asian* music shows moderate representation in research (2,817.73 hours), listenership (part of 58.9 billion views), and population (1.66 billion), but lags in digital catalogs (662 songs).

The representation ratio of *Global North* to *Global South* music, as reflected in *MusicBrainz*, reveals a stark contrast. In the digital realm, this ratio stands at 6.58, while in the ideal space, it drops significantly to 0.66. Additionally, the correlation between the representation of regions in datasets (δ_r) and digital proxies is 0.35 ($\text{Corr}(\pi_{da}, \delta_r)$) and -0.16 ($\text{Corr}(\pi_{dl}, \delta_r)$), while the correlation between the publication based representation (α_r) for regions and *MusicBrainz* is 0.65 ($\text{Corr}(\pi_{da}, \alpha_r)$). This suggests a mild alignment between digital music distribution and the research space but highlights a disconnect between regional viewership and the representation of those regions in music datasets.

Based on this analysis, we assert that while viewership for *European* music and its population is lower compared to *Asian* and *African* regions, *Global North* music datasets disproportionately dominate the research space. In contrast, *South Asian* and *African* music remain significantly underrepresented in music datasets, even though their digital consumption and real-world presence are comparable to, and in some cases exceed, those of *European* and *American* regions.

When compared to the digital space, we observe an interesting contrast with *Hip-Hop* being highly represented with 22562 songs in the *MusicBrainz* database, while having only 6% representation, while *Electronic* music has only 14025 songs but has 13.1% representation. Using *SoundChart* data we observed that *Classical* music has only 13.72 billion views but has 13.5% representation; meanwhile, *Folk*, *Blues*, and *Hip-Hop* have more than 50 billion views each but have

6% or less representation. Among styles represented in the datasets (δ_g), pop is 3 times more popular than *Hip-Hop*, but in the digital music market, the situation is just the reverse of this, where *Hip-Hop* is 3 times more popular than *Pop*. In contrast to the region-wise distribution, the research domain shows a strong correlation with the digital space in genre-specific scenarios, with correlation of 0.71 ($\text{Corr}(\pi_{da}, \rho_s)$) and 0.62 ($\text{Corr}(\pi_{dl}, \rho_s)$) for the genre representation relative to *MusicBrainz* and *SoundCharts* respectively. Genres like *Rock*, *Pop*, and *Electronic* are well represented in digital space as well as in AI research. On the other hand, *Country* and *Classical*, while not popular in digital space, are well-represented in research. *Hip-hop*, *Experimental*, *Folk*, *Blues*, and *Jazz* exhibit the opposite pattern; these genres are popular in the digital space, but not so in research.

Why Does it Matter?

The marginalization of *Global South* music genres in AI-driven music generation has profound implications. Historically, colonization has favored those who control technology while disadvantaging the colonized (contributors 2024). As Lund (2019) notes, Western narratives continue to dominate global pop, with local traditions in Turkey, Brazil, and Peru often gaining recognition only after Western reissues or collectors' interest (e.g., Arşivplak, DJ Tahira²³, Barbès Records²⁴). These cases highlight cultural homogenization, where value is conferred externally rather than internally sustained. Building on this concern, Collins and Grierson (2024) warn that AI could intensify such dynamics by monopolizing music production, potentially erasing diverse traditions in favor of uniform outputs. As AI becomes further embedded in the music industry, the vibrant melodies of the Global Majority risk both erosion and extinction. This trajectory threatens not only the survival of these cultural treasures but also the richness of global musical diversity. In what follows, we examine the consequences of the under-

²³<https://www.djtahira.com/>

²⁴<https://www.barbesbrooklyn.com/>

representation of musical styles in AI-generated music and outline recommendations to mitigate its risks.

- **Bias in Evaluation.** Generative music models are often evaluated using audio embedding backbones such as PANN-CNN14 (Kong et al. 2020) and VGGish²⁵, which map audio into fixed-length embeddings for similarity metrics like Fréchet Audio Distance (FAD) (Kilgour et al. 2018) and KL divergence (Kullback and Leibler 1951). As we saw, global music datasets show strong regional and genre disparities, with several *Global South* regions underrepresented (e.g., Oceania, Central Asia, Africa) contributing < 1% of data and genres like Folk only 2.1%.

Large-scale backbone models, reliant on such datasets, are expected to inherit these gaps, limiting their ability to encode features such as timbres, rhythmic structures, and ornamentations. To better understand these biases, we compare the representation of guitar, piano, drums, globally popular instruments, to sitar, tabla, accordion, bagpipes regional instruments in the training data. As shown in Table 3, for both VGGish and PANN-CNN14, 52.33% and 67.54% of the training clips contain at least one of guitar, piano, drums, whereas the union of sitar, tabla, accordion, bagpipes totals < 3%, and none of them individually has higher than 1%.

Given these instruments’ central role in Hindustani Classical (sitar, tabla) and folk (accordion, bagpipes) traditions, their underrepresentation risks poor embedding quality and biased evaluations against such genres, akin to fairness issues observed in text embeddings (Bolukbasi et al. 2016). We propose a call to action to expand datasets for underrepresented genres and regions and re-train evaluation backbone models on datasets with better representation to ensure fairness (see Definition of fairness and representation).

- **Research and Technological Exclusion.** The exclusion of *Global South* styles from AI training datasets restricts their evolution in the digital age, risking stagnation while others continue to expand. Despite rich traditions, researchers from the *Global South* contribute only 12.1% of publications (see Results section), with most work centered on East Asian, American, & European music (Figure 2).

As shown in Results section, representation gaps vary: Oceania and Latin America show higher research representation than data representation, South Asia & Middle East show the reverse, while Africa & Central Asia remain critically low in both, risking loss of regional traditions. Similarly, for genres Folk, Blues, Country, & Easy Listening remain under 3%, far behind Pop, Electronic, and Classical.

- **Dominance of Symbolic Music Generation.** As a result of these imbalances, symbolic music generation has become the dominant paradigm, accounting for 51% of surveyed works and nearly 90% of ACM publications.

²⁵<https://github.com/tensorflow/models/tree/master/research/audioset/vggish>

Its prevalence stems from its alignment with the structured conventions of *Global North* genres. However, this approach is inadequate in capturing the improvisation, ornamentation, and microtonality that are central to many *Global South* traditions (e.g., Ragas, poly-rhythms, Maqam).

Continued infrastructure development centered on symbolic methods risks further marginalizing these traditions, excluding them from the benefits of future advancements. Further, if left unchecked, such biases risk cultural erosion and deepening economic divides (Okolo 2023; Forum 2023). We urge conferences, specifically ones with specific music focus, including ISMIR, NIME, SMC, ICASSP, EURASIP, and EUSIPCO to address these disparities through diversity-focused workshops, hackathons, and initiatives.

- **Cultural Erosion.** AI models are trained on datasets based on content drawn from digital platforms, where representation is heavily skewed toward the *Global North*. Our initial analysis also reflects this imbalance: the ratio of *Global North* to *Global South* representation is 6.09 in research datasets and increases to 6.58 in the digital music space with $Corr(\pi_{da}, \delta_g) = 0.71$, $Corr(\pi_{da}, \delta_r) = 0.35$, and $Corr(\pi_{da}, \alpha_r) = 0.65$ indicating a strong correlation between the digital space and the dataset/research publication space. By relying primarily on *Global North* music datasets, AI models risk reinforcing existing cultural biases and further marginalizing musicians and composers from the *Global South*, as regions and musical styles that are underrepresented in the training data are less likely to appear in generated outputs.

Furthermore, as AI-generated music scales, it begins to feed back into the digital ecosystem, creating a self-reinforcing loop. As this AI-generated content increasingly dominates digital spaces, existing biases become further entrenched. Hence, the scale of AI music generation and the skew of training data towards digital content dominated by *Global North* datasets (Team 2023) risks reinforcing existing cultural bias and marginalizing *Global South* musicians. This puts traditional music like *Hindustani folk* and *Arabic Maqam* at risk of erosion.

- **The Vicious Loop of Synthetic Music Data.** Synthetic data generation can reinforce and amplify existing biases in music representation. The skew in representation in the larger data is often reflected in the synthetically generated music data. As noted by Agarwal et al. (2021), training on model-generated data can create feedback loops that exacerbate disparities. To illustrate, the *MusicBench* dataset (Melechovsky et al. 2024), which augments *MusicCaps* (Agostinelli et al. 2023), expands from 5k to 37k samples while preserving its skewed genre distribution, boosting the absolute counts of overrepresented genres such as classical (13.7%), electronic (15.6%), and rock (10.5%), and leaving Asian (2.6%), hip-hop (3.4%), and country (5.3%) comparatively scarce. Such scaling risks locking models into biased generation patterns, further hindering fair musical diversity.

Category	Problem	Risks	Mitigation Strategy	Severity
Moderately represented	Inconsistent quality and adherence to the prompt.	Produces music that may not match the prompt, impacting information retrieval.	Improve model architectures and use more descriptive prompts.	Moderate
Under-represented	Instruments and melodies are out of sync, resulting in poor-quality music.	Weak representation leads to less impactful music generation.	Use better datasets and warn users of potential inaccuracies.	High
Unrepresented	Music diverges entirely from the genre.	Misleads listeners, misrepresenting the genre digitally.	Avoid generating samples for unfamiliar genres.	Very High

Table 2: Problems of under-representation of musical genres in AI systems, their risks and mitigation strategies.

Instrument	VGGish (%)	PANN (%)
Guitar	27.39	43.97
Drums	19.14	13.75
Piano	5.80	9.82
Accordion	1.77	2.5
Clarinet	0.24	1.81
Bagpipes	0.20	1.52
Sitar	0.09	1.35
Tabla	0.08	1.47
Bassoon	0.08	—

Table 3: Percentage Composition of Instruments in the Collection for VGGish and PANN Models

Conclusion and Recommendation

In this article, we made an attempt to bring out an important gap in AI-driven music generation: the noticeable lack of representation of music genres from the *Global South*. While AI tools have made incredible progress in generating music, they overwhelmingly focus on the musical styles from the *Global North*, leaving out many unique and culturally rich styles from regions such as Africa, Latin America, South Asia, and the Middle East. Global musical cultures evolve through economic, technological, and community-driven efforts to decolonize and diversify (Tan 2021). In AI-generated music, fully egalitarian outcomes may be impossible (Choudhury and Deshpande 2021), yet concrete strategies can still foster broader inclusion of diverse traditions. Here are some concrete steps that the AI-music research community can adopt to ensure that future AI-music generation systems can serve everyone on the planet more equitably.

- **Acknowledging Model Limitations:** The solution to a problem begins by acknowledging it. Therefore, music generation models as well as research papers should explicitly identify and acknowledge the limitations of their modeling systems in their research. For instance, symbolic music generation often falls short in capturing microtonal variations (e.g., Shrutis²⁶ in Indian Classical)

²⁶The shruti is the smallest interval of pitch that the human ear

and intricate irregular rhythmic structures (Shah 2001). It lacks the flexibility to convey the nuanced elements present in genres like Indian Classical or Maqamat music. Clearly stating these limitations within the paper will encourage the research community to address and improve upon the constraints of these models.

- **Avoid Generation When Uncertain:** Models will inevitably encounter genres they are under-trained on or cannot generate confidently, since an ideal egalitarian approach is difficult to achieve practically. In such cases, issue warnings for underrepresented genres and avoid generating for unrepresented ones to prevent misleading outputs and distortion of the digital music space. Table 2 summarises the issues, risks, and our recommendations.
- **Explicit Mention of the Genres:** Many datasets as well as research papers do not describe the genres explicitly. This makes it difficult to identify and address the gaps. Much like the Bender Rule in NLP (Bender and Friedman 2018), which advocates for clear articulation of the languages in data statements for training, tuning, or testing a system, music generation papers should clearly state the genres used for model training and evaluation. This ensures accurate claims about model universality and prevents misinterpretations, especially when certain genres may be underrepresented. Including genre information in the abstract or main body will promote clarity regarding the model’s applicability across musical styles.
- **Investing in Inclusive Datasets:** Initiatives like the Masakhane Project (Orife et al. 2020) in NLP show how community-driven, open-source efforts can build representative datasets. A similar approach in music could ensure diverse regional and genre coverage, addressing gaps such as *Hip-Hop*, which is digitally prominent but underrepresented in research. Following examples like Google’s Inclusive Images Competition (Atwood et al. 2020), AI music generation should balance datasets by both genre and region (Bryan-Kinns and Li 2024).
- **Transfer Learning for Underrepresented Styles:** Similar to the language domain, where transfer learning is used for low-resource languages (Joshi et al. 2024; Zoph

can detect and a singer or musical instrument can produce.

et al. 2016; Lankford, Afli, and Way 2023), the music research community should invest in techniques for sample-efficient cross-genre transfer of musical styles, instruments, melodic structures, and so on (Doosti and Guzdial 2023).

- **Inclusive Evaluation:** Genre-specific evaluations are rarely conducted or clearly reported, leaving the performance of many styles unknown. Prior work (Xiong et al. 2023; Yang and Lerch 2018) highlights the need for subjective assessments, while automatic evaluations require backbone models finetuned on diverse music to avoid misrepresentation (See Discussion section). Researchers should also develop metrics that generalize across genres without repeated finetuning.

References

- Agarwal, K.; Zhao, J.; Wang, D.; Zhao, K.-W.; and Chang, K.-W. 2021. Fairness Feedback Loops: Training on Synthetic Data Amplifies Bias. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.
- Agostinelli, A.; Denk, T. I.; Borsos, Z.; Engel, J.; Verzett, M.; Caillon, A.; Huang, Q.; Jansen, A.; Roberts, A.; Tagliasacchi, M.; et al. 2023. Musiclm: Generating music from text. *arXiv preprint arXiv:2301.11325*.
- Atwood, J.; Halpern, Y.; Baljekar, P.; Breck, E.; Sculley, D.; Ostyak, P.; Nikolenko, S. I.; Ivanov, I.; Solovyev, R.; Wang, W.; et al. 2020. The inclusive images competition. In *The NeurIPS'18 Competition: From Machine Learning to Intelligent Conversations*, 155–186. Springer.
- Bender, E. M.; and Friedman, B. 2018. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6: 587–604.
- Bolukbasi, T.; Chang, K.-W.; Zou, J.; Saligrama, V.; and Kalai, A. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16*, 4356–4364. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781510838819.
- Bryan-Kinns, N.; and Li, Z. 2024. Reducing Barriers to the Use of Marginalised Music Genres in AI. *arXiv preprint arXiv:2407.13439*.
- Cholewiak, S. A.; Ipeirotis, P.; Silva, V.; and Kannawadi, A. 2021. SCHOLARLY: Simple access to Google Scholar authors and citation using Python.
- Choudhury, M.; and Deshpande, A. 2021. How Linguistically Fair Are Multilingual Pre-Trained Language Models? *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(14): 12710–12718.
- Collins, N.; and Grierson, M. 2024. Avoiding an AI-imposed Taylor's Version of all music history. *arXiv preprint arXiv:2402.14589*.
- contributors, W. 2024. Technopoly.
- Copet, J.; Kreuk, F.; Gat, I.; Remez, T.; Kant, D.; Synnaeve, G.; Adi, Y.; and Défossez, A. 2024. Simple and controllable music generation. *Advances in Neural Information Processing Systems*, 36.
- Dhariwal, P.; Jun, H.; Payne, C.; Kim, J. W.; Radford, A.; and Sutskever, I. 2020. Jukebox: A generative model for music. *arXiv preprint arXiv:2005.00341*.
- Doosti, A.; and Guzdial, M. 2023. Transfer Learning for Underrepresented Music Generation. *arXiv:2306.00281*.
- Dysart, J. 2018. Give Me That AI Rock and Roll.
- Florida, R.; Mellander, C.; and Stolarick, K. 2010. Music Scenes to Music Clusters: The Economic Geography of Music in the US, 1970–2000. *Environment and Planning A Economy and Space*, 42(4): 785–804.
- Forum, W. E. 2023. The 'AI Divide' Between the Global North and Global South. *World Economic Forum*.
- Henning, S.; Beluch, W.; Fraser, A.; and Friedrich, A. 2023. A Survey of Methods for Addressing Class Imbalance in Deep-Learning Based Natural Language Processing. In Vlachos, A.; and Augenstein, I., eds., *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, 523–540. Dubrovnik, Croatia: Association for Computational Linguistics.
- Joshi, P.; Santy, S.; Budhiraja, A.; Bali, K.; and Choudhury, M. 2020. The state and fate of linguistic diversity and inclusion in the NLP world. *arXiv preprint arXiv:2004.09095*.
- Joshi, S.; Khan, M. S.; Dafe, A.; Singh, K.; Zope, V.; and Jhamtani, T. 2024. Fine Tuning LLMs for Low Resource Languages. In *2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN)*, 511–519.
- Kilgour, K.; Zuluaga, M.; Roblek, D.; and Sharifi, M. 2018. Fr\`echet audio distance: A metric for evaluating music enhancement algorithms. *arXiv preprint arXiv:1812.08466*.
- Kong, Q.; Cao, Y.; Iqbal, T.; Wang, Y.; Wang, W.; and Plumbley, M. D. 2020. Panns: Large-scale pretrained audio neural networks for audio pattern recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28: 2880–2894.
- Kullback, S.; and Leibler, R. A. 1951. On information and sufficiency. *The annals of mathematical statistics*, 22(1): 79–86.
- Lankford, S.; Afli, H.; and Way, A. 2023. adaptM-LLM: Fine-Tuning Multilingual Language Models on Low-Resource Languages with Integrated LLM Playgrounds. *Information*, 14(12).
- Lund, H. 2019. Decolonizing Pop Music.
- Melechovsky, J.; Guo, Z.; Ghosal, D.; Majumder, N.; Herremans, D.; and Poria, S. 2024. Mustango: Toward Controllable Text-to-Music Generation. In Duh, K.; Gomez, H.; and Bethard, S., eds., *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, 8293–8316. Mexico City, Mexico: Association for Computational Linguistics.
- Okolo, C. T. 2023. AI in the Global South: Opportunities and challenges towards more inclusive governance. *Brookings Institution Press*. <https://www.brookings.edu/articles/ai-in-the-global-south-opportunities-and-challengestowards-more-inclusive-governance>.

- Orife, I.; Kreutzer, J.; Sibanda, B.; Whitenack, D.; Siminyu, K.; Martinus, L.; Ali, J. T.; Abbott, J.; Marivate, V.; Kabongo, S.; et al. 2020. Masakhane–Machine Translation For Africa. *arXiv preprint arXiv:2003.11529*.
- Rawls, J. 1958. Justice as fairness. *The philosophical review*, 67(2): 164–194.
- Savage, P. E. 2019. Cultural evolution of music. *Palgrave Communications*, 5(1): 1–12.
- Schneider, F.; Kamal, O.; Jin, Z.; and Schölkopf, B. 2024. Moûsai: Efficient text-to-music diffusion models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 8050–8068.
- Shah, N. 2001. *Visualization of Indian Classical Music*. Ph.D. thesis, Citeseer.
- Tan, S. E. 2021. Special issue: Decolonising music and music studies. *Ethnomusicology Forum*, 30(1): 4–8.
- Tao, Y.; Viberg, O.; Baker, R. S.; and Kizilcec, R. F. 2024. Cultural bias and cultural alignment of large language models. *PNAS Nexus*, 3(9): pgae346.
- Team, I. M. A. 2023. Music and AI: The Pros, Cons, and Ethical Implications. *Intercontinental Music Awards*.
- Ungruh, R.; Dinnissen, K.; Volk, A.; Pera, M. S.; and Hauptmann, H. 2024. Putting Popularity Bias Mitigation to the Test: A User-Centric Evaluation in Music Recommenders. In *Proceedings of the 18th ACM Conference on Recommender Systems*, RecSys ’24, 169–178. New York, NY, USA: Association for Computing Machinery. ISBN 9798400705052.
- Xiong, Z.; Wang, W.; Yu, J.; Lin, Y.; and Wang, Z. 2023. A Comprehensive Survey for Evaluation Methodologies of AI-Generated Music. *arXiv preprint arXiv:2308.13736*.
- Yang, L.-C.; and Lerch, A. 2018. On the evaluation of generative models in music. *Neural Computing and Applications*, 32(9): 4773–4784.
- Zoph, B.; Yuret, D.; May, J.; and Knight, K. 2016. Transfer learning for low-resource neural machine translation. *arXiv preprint arXiv:1604.02201*.