AI Music Evaluation Metrics

21 Mar 2025

Jon Chun

PROMPT:

Recent (as of March 2025) survey or papers on metrics, benchmarks, leaderboards to asses the quality of AI generated music

PR

recent survey and research articles focused on AI generated art (music, image, text/stories, etc.), creativity, aesthetics as well as related issues like IP, disemployment, economics, etc as secondary add ons

[Grok 3]

### PROMPT #1:

### 

PROMPT #2:

### Key Points

* Research suggests AI-generated art, including music, images, and text/stories, is rapidly evolving, with mixed public and artist perceptions on creativity and aesthetics.
* It seems likely that metrics for quality include **listener indistinguishability** (82% for music) and **public enjoyment** (56% for art), but controversy exists over whether AI art is "**art**" (76% of Americans say no).
* The evidence leans toward significant **market growth**, with the global AI market projected at $826 billion by 2030 and music AI at $38.71 billion by 2033, raising economic and disemployment concerns.
* IP issues are contentious, with 95% of **music creators** demanding **transparency** and 89% of artists worried about outdated **copyright** laws.

### AI-Generated Art Overview

AI-generated art spans music, images, and text/stories, with recent surveys highlighting its impact on **creativity and aesthetics**. Public perception is mixed, with 56% **enjoying** AI art but 76% not considering it "**art**" ([AI Art Statistics 2024](https://artsmart.ai/blog/ai-art-statistics/)). Artists are increasingly using AI for **brainstorming** (65%), but concerns about **ethics and originality** persist, with 74% finding AI artwork unethical ([AI Art Statistics 2024](https://artsmart.ai/blog/ai-art-statistics/)).

### Market and Economic Impacts

The market for AI art is booming, with the **global AI market** valued at **$184 billion in 2024**, expected to reach **$826 billion by 2030**, potentially boosting **GDP by 26%** ([77 AI Statistics & Trends to Quote in 2025](https://planable.io/blog/ai-statistics/)). For music, the market was **$2.9 billion in 2024**, projected to hit $38.71 billion by 2033, with a 17.2% revenue increase by 2025 ([AI in Music Industry Statistics 2025](https://artsmart.ai/blog/ai-in-music-industry-statistics/)). However, 71% of music **creators fear livelihood** threats, predicting a 27% revenue drop by 2028.

### Intellectual Property and Disemployment

IP is a major concern, with 95% of music creators demanding transparency in AI use of copyrighted works and 89% of artists worried about outdated laws ([AI in Music Industry Statistics 2025](https://artsmart.ai/blog/ai-in-music-industry-statistics/); [AI Art Statistics 2024](https://artsmart.ai/blog/ai-art-statistics/)). **Disemployment** fears are significant, with 20% of marketers feeling AI poses a definite threat to jobs and 71% of music creators worried about income loss ([77 AI Statistics & Trends to Quote in 2025](https://planable.io/blog/ai-statistics/); [AI in Music Industry Statistics 2025](https://artsmart.ai/blog/ai-in-music-industry-statistics/)).

### Unexpected Detail: AI in Music Aesthetics

An unexpected finding is that **82% of listeners cannot distinguish AI-generated** music from human compositions, challenging traditional notions of musical authenticity ([AI in Music Industry Statistics 2025](https://artsmart.ai/blog/ai-in-music-industry-statistics/)).

### Comprehensive Analysis of Recent Surveys and Research on AI-Generated Art, Creativity, Aesthetics, and Related Issues

This report provides a detailed examination of recent surveys and research articles focused on AI-generated art, including music, images, text/stories, creativity, aesthetics, and secondary issues such as **intellectual property (IP), disemployment, and economics**. The analysis is based on the most recent and reliable sources available as of March 2025, reflecting the state of the field and its implications for artists, creators, and society.

#### Introduction

AI-generated art has become a **transformative** force across creative industries, with applications in **music, visual arts, and text/stories**. Recent surveys and research highlight its impact on **creativity and aesthetics**, while also raising concerns about IP, disemployment, and economic shifts. This report synthesizes findings from various sources to provide a **holistic view**, addressing the user's query comprehensively.

#### AI-Generated Art: General Findings

AI-generated art encompasses a broad range of mediums, with public perception and artist usage showing both **enthusiasm and skepticism**. A survey from April 2024 found that 27% of Americans have seen AI-generated art, with 56% enjoying it, but 76% do not believe it should be considered "art" ([AI Art Statistics 2024](https://artsmart.ai/blog/ai-art-statistics/)). This reflects a **generational divide**, with 48% of Millennials supporting AI art as art, compared to 61% of the general U.S. adult population disagreeing ([Statista, 2024](https://artsmart.ai/blog/ai-art-statistics/)).

Artist usage is significant, with 65% using text-to-image AI for brainstorming and 36% believing it accurately reflects their artistry ([Academy of Animated Art, 2024](https://artsmart.ai/blog/ai-art-statistics/)). However, ethical concerns are prevalent, with 74% of artists considering AI artwork unethical and 89% worried about outdated copyright laws ([Academy of Animated Art, 2024](https://artsmart.ai/blog/ai-art-statistics/)). Market trends show AI art fetching high values, with the highest traditional AI art selling for $432,000 and AI NFTs reaching $1.1 million, with prices on OpenSea ranging from $244 to $1,631 ([Academy of Animated Art, 2024](https://artsmart.ai/blog/ai-art-statistics/)).

The **global generative AI market** is projected to surpass $1.3 trillion by 2032, with the U.S. market expected to reach $20.29 billion by 2025, driven by sectors like **entertainment and marketing** ([Statista, 2024](https://artsmart.ai/blog/ai-art-statistics/)). Social implications include concerns about **deepfakes**, with nearly 70% of global respondents worried, and potential **biases amplified** by AI ([Statista, 2024](https://artsmart.ai/blog/ai-art-statistics/)).

#### AI-Generated Music: Detailed Insights

Research on AI-generated music highlights significant market growth and adoption. The global AI in music market was valued at $2.9 billion in 2024, expected to reach $38.71 billion by 2033, with a compound annual growth rate (CAGR) of 25.8% from 2024-2033 ([Market.us, 2024](https://artsmart.ai/blog/ai-in-music-industry-statistics/)). Revenue in the music industry is expected to boost by 17.2% by 2025, driven by AI-generated music, with cloud-based solutions dominating 71.4% of the market share ([Market.us, 2024](https://artsmart.ai/blog/ai-in-music-industry-statistics/)).

**Adoption rates vary by genre,** with electronic music leading at 54%, followed by hip-hop at 53%, and advertising music at 52%, while **traditional/world music lags** at 30% ([Market.us, 2024](https://artsmart.ai/blog/ai-in-music-industry-statistics/)). A survey of 1,533 music producers found that **36.8% already use AI tools** in their workflow, with **60% of musicians adopting AI overall**, including 30.6% for **mastering** and 38% for **artwork** ([Ditto Music, 2023](https://artsmart.ai/blog/ai-in-music-industry-statistics/); [Rare Connections, 2024](https://artsmart.ai/blog/ai-in-music-industry-statistics/)). Additionally, 74% of internet users use AI for music **discovery**, indicating its integration into consumer behavior ([Rare Connections, 2024](https://artsmart.ai/blog/ai-in-music-industry-statistics/)).

On creativity and aesthetics, a striking finding is that **82% of listeners cannot distinguish** AI-generated music from human compositions, suggesting **high aesthetic** quality ([Market.us, 2024](https://artsmart.ai/blog/ai-in-music-industry-statistics/)). However, 54% of surveyed creators agree that AI can **assist the human creative process**, with Australian and New Zealand songwriters being early adopters ([APRA AMCOS, 2024](https://www.apraamcos.com.au/about-us/news-and-events/ai-in-music-report)).

IP concerns are significant, with 95% of artists demanding transparency in AI's use of copyrighted works and 90% believing AI companies should ask permission ([Market.us, 2024](https://artsmart.ai/blog/ai-in-music-industry-statistics/)). A study found that **21% of creators have faced copyright strikes**, with 21% facing multiple strikes, and potential revenue loss estimated at €950 million by 2028 in France and Germany ([Music Business Worldwide, 2024](https://www.musicbusinessworldwide.com/generative-ai-music-market-to-be-worth-3bn-by-2028-says-report-commissioned-by-sacem-and-gema/)).

**Disemployment fears** are pronounced, with **71% of music creators** concerned that AI could lead to them no longer being able to make a living, and a Goldmedia study warning of a **27% revenue drop** by 2028 ([Market.us, 2024](https://artsmart.ai/blog/ai-in-music-industry-statistics/)). Economically, **$50 billion has been invested in AI in Europe** since the launch of ChatGPT, with companies like Amper Music raising $4 million, and a predicted €2.7 billion market gap highlighting economic shifts ([Market.us, 2024](https://artsmart.ai/blog/ai-in-music-industry-statistics/)).

#### AI-Generated Images: Market and Perception

For **AI-generated images**, public awareness is growing, with **54% able to tell the difference** between AI-generated and human-made images, but concerns about misinformation are high, with 53% of Americans worried about fake news spread by AI images ([Academy of Animated Art, 2024](https://artsmart.ai/blog/ai-art-statistics/)). Artist concerns include 89% worrying about outdated copyright laws and 74% finding AI artwork unethical, with 73% demanding permission before their work is used to train AI ([Academy of Animated Art, 2024](https://artsmart.ai/blog/ai-art-statistics/)).

Market trends show the AI image generator market valued at $336.3 million in 2023, expected to grow at a **CAGR of 17.5% from 2024-2032**, reflecting increasing adoption ([Global Market Insights, 2024](https://artsmart.ai/blog/ai-art-statistics/)). **Influencer usage** is notable, with 51.8% using Canva’s AI features and 36.4% using Photoshop’s AI, and 67.1% disclosing AI usage to followers ([Statista, 2024](https://artsmart.ai/blog/ai-art-statistics/)).

#### AI-Generated Text/Stories: Usage and Impact

For text and stories, **29% of marketers use AI for content creation**, with **42% using it for social media caption**s and **28% for images**, reflecting its integration into **marketing workflows** ([Influencer Marketing Hub, 2024](https://planable.io/blog/ai-statistics/)). **Trust varies**, with 30% trusting AI, 15% distrusting it, and 48% saying it depends on the tool ([Statista, 2024](https://planable.io/blog/ai-statistics/)). A survey found that 70.6% of marketers believe **AI can outperform humans in key marketing tasks**, with 34.1% reporting significant improvements in outcomes ([Influencer Marketing Hub, 2024](https://planable.io/blog/ai-statistics/)).

IP concerns include 35% of marketers citing **privacy and data security** as primary issues, reflecting legal challenges ([Planable Survey, 2024](https://planable.io/blog/ai-statistics/)). Disemployment fears are evident, with 50% not seeing AI as a threat, but 21% unsure, 18% feeling threatened, and 20% believing it poses a definite threat ([Planable Survey, 2024](https://planable.io/blog/ai-statistics/)). Economically, the global AI market is valued at $184 billion in 2024, expected to reach $826 billion by 2030, with a potential $4.4 trillion value across industries and a 26% GDP boost by 2030 ([Statista, 2024](https://planable.io/blog/ai-statistics/)).

#### Secondary Issues: IP, Disemployment, and Economics

**IP remains a contentious** issue, with 95% of music creators demanding transparency and 90% believing AI companies should ask permission, while artists sue for unauthorized use of copyrighted data ([Market.us, 2024](https://artsmart.ai/blog/ai-in-music-industry-statistics/)). The **U.S. Copyright Office's 2025 report reaffirms human** authorship as central to copyright, but AI-generated content raises complex legal questions ([Center for Art Law, 2025](https://itsartlaw.org/2025/03/04/recent-developments-in-ai-art-copyright-copyright-office-report-new-registrations/)).

Disemployment fears are significant, with 71% of music creators worried about livelihood threats and 20% of marketers feeling AI poses a definite job threat, predicting revenue drops and job shifts ([Market.us, 2024](https://artsmart.ai/blog/ai-in-music-industry-statistics/); [Planable Survey, 2024](https://planable.io/blog/ai-statistics/)). Economically, the rapid growth of AI markets, with $25.2 billion invested in generative AI in 2023 (9x 2022), and 92% of large companies reporting ROI, underscores its economic impact ([IBM, 2023](https://www2.multivu.com/players/English/9240059-ibm-2023-global-ai-adoption-index-report/); [Statista, 2024](https://planable.io/blog/ai-statistics/)).

#### Creativity and Aesthetics: Detailed Analysis

Creativity is enhanced by AI, with 65% of artists using it for brainstorming and 70.6% of marketers believing it outperforms humans in tasks ([Academy of Animated Art, 2024](https://artsmart.ai/blog/ai-art-statistics/); [Influencer Marketing Hub, 2024](https://planable.io/blog/ai-statistics/)). However, concerns about originality are notable, with 28% of artists worried about AI-generated works' lack of originality ([YouGov, 2024](https://artsmart.ai/blog/ai-art-statistics/)).

Aesthetics are challenged by AI's ability to produce indistinguishable outputs, with 82% of listeners unable to differentiate AI music from human compositions and 54% able to identify AI images, reflecting both acceptance and scrutiny ([Market.us, 2024](https://artsmart.ai/blog/ai-in-music-industry-statistics/); [Academy of Animated Art, 2024](https://artsmart.ai/blog/ai-art-statistics/)). Public perception remains mixed, with 76% of Americans not considering AI art as "art," highlighting a cultural debate ([Academy of Animated Art, 2024](https://artsmart.ai/blog/ai-art-statistics/)).

#### Tables: Key Statistics

Below are tables summarizing key findings across categories:

**Table 1: Public Perception and Adoption Rates**

| **Category** | **Statistic** | **Source** |
| --- | --- | --- |
| Seen AI Art | 27% of Americans | [AI Art Statistics 2024](https://artsmart.ai/blog/ai-art-statistics/) |
| Enjoy AI Art | 56% of viewers | [AI Art Statistics 2024](https://artsmart.ai/blog/ai-art-statistics/) |
| Believe AI is Art | 48% Millennials, 61% disagree overall | [Statista, 2024](https://artsmart.ai/blog/ai-art-statistics/) |
| Music Producers Using AI | 36.8% | [Ditto Music, 2023](https://artsmart.ai/blog/ai-in-music-industry-statistics/) |
| Marketers Using AI for Text | 29% | [Influencer Marketing Hub, 2024](https://planable.io/blog/ai-statistics/) |

**Table 2: Market and Economic Impacts**

| **Category** | **Statistic** | **Source** |
| --- | --- | --- |
| Global AI Market 2024 | $184 billion | [Statista, 2024](https://planable.io/blog/ai-statistics/) |
| Projected by 2030 | $826 billion, 26% GDP boost | [Statista, 2024](https://planable.io/blog/ai-statistics/) |
| AI Music Market 2024 | $2.9 billion | [Market.us, 2024](https://artsmart.ai/blog/ai-in-music-industry-statistics/) |
| Projected by 2033 | $38.71 billion, 25.8% CAGR | [Market.us, 2024](https://artsmart.ai/blog/ai-in-music-industry-statistics/) |
| AI Image Market 2023 | $336.3 million, 17.5% CAGR (2024-2032) | [Global Market Insights, 2024](https://artsmart.ai/blog/ai-art-statistics/) |

#### Conclusion

As of March 2025, AI-generated art is reshaping creative industries, with significant advancements in music, images, and text/stories. While it enhances creativity and aesthetics, it also raises complex issues around IP, disemployment, and economics, necessitating ongoing research and policy development to balance innovation with ethical considerations.

#### Key Citations

* [AI Art Statistics 2024: Explore 25+ Essential AI Art Statistics](https://artsmart.ai/blog/ai-art-statistics/)
* [AI in Music Industry Statistics 2025: Market Growth & Trends](https://artsmart.ai/blog/ai-in-music-industry-statistics/)
* [77 AI Statistics & Trends to Quote in 2025 + Own Survey Results](https://planable.io/blog/ai-statistics/)
* [Recent Developments in AI, Art & Copyright: Copyright Office Report](https://itsartlaw.org/2025/03/04/recent-developments-in-ai-art-copyright-copyright-office-report-new-registrations/)
* [AI as an Artist? A Two-Wave Survey Study on Attitudes Toward Using AI in Art](https://www.sciencedirect.com/science/article/pii/S0304422X23000797)
* [4 Ways A.I. Impacted the Art Industry in 2024 | Artnet News](https://news.artnet.com/art-world/a-i-art-industry-2024-2591678)

[Perplexity]

# Recent Advances in Evaluation Methods for AI-Generated Music: A 2025 Perspective

As of March 2025, artificial intelligence has transformed music creation, with AI-generated content becoming increasingly sophisticated and widespread. This transformation has necessitated the development of robust frameworks to evaluate the quality of AI-generated music. This report examines recent surveys, papers, and benchmarks that assess AI music generation quality, highlighting key metrics, evaluation methodologies, and industry standards.

## Emerging Evaluation Frameworks and Metrics

### Academic Metrics Development

Recent academic work has significantly advanced our understanding of how to evaluate AI-generated music. A notable contribution comes from University of Padova researcher Davide Bertini, whose 2024/2025 study focused on evaluating **fine-tuned versions of Meta's MusicGEN Transformer** model[1]. This research employed three primary metrics that have gained traction in the AI music evaluation community:

* **Fréchet Audio Distance (FAD)**: Measures the similarity between distributions of real and generated audio features
* **Kullback-Leibler Divergence**: Quantifies the difference between probability distributions of generated and reference music
* **Contrastive Language-Audio Pretraining (CLAP) score**: Evaluates the alignment between generated music and textual descriptions[1]

These metrics offer objective ways to compare AI-generated music against reference sets like MusicCaps and specialized databases, helping researchers understand how model refinements affect output quality.

### Comprehensive Evaluation Methodologies

A landmark development in the field has been the publication of "**A Comprehensive Survey for Evaluation Methodologies of AI-Generated Music**," which systematically analyzes subjective, objective, and combined assessment approaches[7]. This survey highlights that while objective metrics provide reproducible results, they often lack musical interpretability, leading researchers to employ multi-faceted evaluation strategies.

The survey emphasizes three key evaluation paradigms:

1. Subjective evaluation through user studies
2. Objective evaluation using computational metrics
3. Hybrid approaches that combine human judgment with algorithmic assessment[7]

## Benchmark Frameworks and Standardization

### MARBLE: Setting Industry Standards

The **Music Audio Representation Benchmark for universaL Evaluation (MARBLE)** has emerged as a significant benchmark system for evaluating music-related AI capabilities. Though initially introduced in late 2023, MARBLE has become increasingly influential in 2025, offering a comprehensive taxonomy with **four hierarchy levels: acoustic, performance, score, and high-level description**[2][10].

MARBLE establishes a unified protocol based on multiple tasks across various publicly available datasets, providing standardized assessment of all open-source pre-trained models developed for music recordings[10]. Its framework has been instrumental in allowing fair comparisons between different approaches to music AI.

### Bridge.audio's Comparative Benchmark

In **January 2025, Bridge.audio** published a detailed benchmark comparing several leading AI systems for music analysis: **Bridge.audio, Musiio, Cyanite, and Aims**[4]. While focused primarily on **analysis rather than generation**, this benchmark evaluated these systems across diverse songs, including recent hits by Kendrick Lamar, Billie Eilish, Chappell Roan, and Bruno Mars & Lady Gaga[4]. The evaluation assessed capabilities in **genre classification, mood detection, key identification, BPM detection, and theme recognition,** establishing a comparative framework that has influenced generation evaluation practices as well.

## Competitions and Leaderboards

### 2025 Automatic Music Transcription Challenge

A significant upcoming event in the field is the **2025 Automatic Music Transcription (AMT) Challeng**e, scheduled to run from April to May 2025[15]. This competition requires participants to submit programs that convert audio recordings of **classical music into MIDI** format, with evaluation based on multiple factors:

* Accuracy of instrument identification
* Pitch detection precision
* Onset and offset timing
* Dynamic expression capture[15]

The competition will primarily use synthesized audio with a small percentage of human performances, providing a controlled environment for evaluating AI transcription capabilities that directly relates to generation quality assessment.

### AI Leaderboard Developments

While not specific to music, LM Arena has emerged as an influential AI evaluation platform in early 2025[17]. Its user-driven, blind assessment methodology—where users evaluate outputs without knowing which model produced them—has inspired similar approaches in music AI evaluation. This methodology addresses concerns about **benchmark gaming** that has plagued traditional leaderboards[17].

## Consumer Perception and Industry Impact

### Public Reception of AI-Generated Music

A **September 2024 survey by Authority Hacke**r offers valuable insights into public perception of AI-generated music, finding that while 60.58% of respondents are **willing to listen** to AI-generated music, an overwhelming 93.75% **do not value it equally** to human-created content[6]. This perception gap presents a challenge for evaluation frameworks that must account for **both technical quality and listener reception.**

The survey also revealed that 57.67% of respondents **suspect they have unknowingly listened** to AI-produced music, highlighting the importance of transparency in evaluation frameworks[6]. Additionally, 90% of respondents **demanded clear labeling** of AI-generated music, suggesting that **attribution and transparency** are becoming essential components of quality assessment.

### Market and Industry Perspectives

The rapidly growing AI music market, projected to reach $6.2 billion in 2025 and expand to $38.7 billion by 2033[12], has intensified the need for reliable evaluation metrics. Industry sources indicate that approximately 60% of musicians now use AI for tasks including mastering and composition, with **82% of listeners reportedly unable to distinguish** between human and AI-created music **in certain contexts**[12].

A January 2025 assessment by Joe Pater notes that AI music generation has reached a level where it can produce content "indistinguishable from non-AI music, at least in some genres, and when used correctly"[11]. This convergence of AI and human-created music quality underscores the increasing sophistication needed in evaluation methodologies.

## Conclusion

The evaluation of AI-generated music has evolved significantly by March 2025, with metrics becoming more sophisticated and evaluation frameworks more comprehensive. The field has moved beyond simple technical assessments to incorporate elements of human perception, artistic quality, and contextual appropriateness.

Key developments include the **standardization of objective metrics** like FAD and CLAP scores, the establishment of **comprehensive benchmarks** like MARBLE, and increasing attention to **subjective evaluation** methodologies that capture human responses to AI-created music. As AI music generation continues to advance, evaluation frameworks will likely continue to evolve, incorporating more nuanced understanding of musical quality and artistic value.

The **tension between technical excellence and human acceptance** remains a central challenge, suggesting that future evaluation approaches will need to balance objective metrics with subjective assessment to fully capture the quality of AI-generated music.

Citations: [1] <https://thesis.unipd.it/retrieve/2c1c2450-80e9-4614-81b2-461125cd3721/BERTINI_DAVIDE.pdf> [2] <https://proceedings.neurips.cc/paper_files/paper/2023/hash/7cbeec46f979618beafb4f46d8f39f36-Abstract-Datasets_and_Benchmarks.html> [3] <https://www.linkedin.com/pulse/when-brands-embrace-gen-ai-music-soundout-music-testing-1kp5e> [4] <https://www.bridge.audio/blog/benchmark-of-the-best-ai-for-music-analysis-in-2025/> [5] <https://www.soundverse.ai/blog/article/top-5-ai-music-generators-of-2025> [6] <https://www.authorityhacker.com/ai-music-survey/> [7] <https://www.zeyuxiong.com/publication/aimusiceva/> [8] <https://dasarpai.com/dsblog/exploring-ai-benchmarks-and-leaderboards> [9] <https://www.nature.com/articles/s41598-025-92327-8> [10] <https://arxiv.org/abs/2306.10548> [11] <https://websites.umass.edu/pater/2025/01/06/ai-music-generation-is-now-really-really-good/> [12] <https://simplebeen.com/ai-music-statistics/> [13] <https://www.anrfactory.com/10-ways-ai-is-likely-to-shape-the-music-industry-in-2025/> [14] <https://americanspcc.org/top-5-ai-music-generators-of-2025-revolutionizing-sound-creation-and-inspiring-creativity/> [15] <https://ai4musicians.org/transcription/2025transcription.html> [16] <https://www.paperdigest.org/2020/07/recent-papers-on-ai-for-music/> [17] <https://www.christopherspenn.com/2025/02/almost-timely-news-%F0%9F%97%9E%EF%B8%8F-the-state-of-the-state-of-the-art-of-ai-2025-02-02/> [18] <https://www.thebusinessresearchcompany.com/report/generative-artificial-intelligence-ai-in-music-global-market-report> [19] <https://stewarttownsend.com/the-future-of-ai-music-in-q2-2025-what-to-expect/> [20] <https://studio.gaudiolab.io/blog/39> [21] <https://www.audiocipher.com/post/ai-music-app> [22] <https://artsmart.ai/blog/ai-in-music-industry-statistics/>

Answer from Perplexity: pplx.ai/share

[Gemini 3]

# **Evaluating the Quality of AI-Generated Music: Metrics, Benchmarks, and Future Directions**

**1. Introduction: The Rise of AI in Music and the Imperative for Robust Evaluation**

The field of artificial intelligence has witnessed remarkable advancements in recent years, extending its capabilities into creative domains such as music generation. Contemporary AI models can now produce musical compositions with increasing sophistication, blurring the lines between human and machine creativity 1. This progress is underscored by the emergence of a new generation of generative models in 2023, which have amplified both the potential and the challenges associated with AI-generated music 2. The market for AI in music is experiencing substantial growth, with projections estimating a value of $38.71 billion by 2033, reflecting the increasing integration of AI into the music industry 3. This integration is further evidenced by the growing adoption of AI tools among musicians and producers; statistics from 2024 indicate that 60% of musicians utilize AI in their projects, and over a third of professional music producers regularly employ these tools 4. Notably, a significant portion of listeners, around 82%, report difficulty in distinguishing between music created by AI and that composed by humans 4. This convergence of technological progress and widespread adoption highlights the critical need for robust and reliable methods to evaluate the quality of AI-generated music.

The development of effective evaluation tools, particularly objective metrics, has proven to be a significant challenge in the realm of music generation 5. While objective metrics offer the advantage of quantifiable analysis, they often fall short in capturing the nuances of musical quality as perceived by humans 6. Consequently, researchers frequently resort to subjective user studies to assess the quality of AI-generated music. However, these studies can be resource-intensive and may lack the reproducibility offered by objective measures 6. The inherent subjectivity of music appreciation further complicates the establishment of universally accepted objective metrics, creating a fundamental hurdle in the field 6. The difficulty in distinguishing AI-generated music from human compositions 4 underscores the necessity for evaluation methods that extend beyond mere detection and delve into aspects such as musicality, creativity, and aesthetic appeal. This report aims to provide a comprehensive overview of recent research on the evaluation of AI-generated music, focusing on the metrics, benchmarks, and leaderboards employed. It will specifically address the challenges associated with evaluating creativity, aesthetics, and out-of-distribution music generated by AI, while also assessing the comprehensiveness of current evaluation approaches and identifying existing gaps and limitations in this rapidly evolving field.

**2. Subjective Metrics for Evaluating AI-Generated Music**

Given the intrinsic subjective nature of music, human-centric evaluation methods play a crucial role in assessing the quality of AI-generated compositions. Music listening tests represent the most common approach in subjective evaluation 6. These tests often employ methodologies such as the Musical Turing Test, which assesses whether listeners can distinguish between music generated by a machine and that created by a human. Additionally, subjective query metrics are utilized, frequently employing Likert scales to quantify listener preferences across various parameters, including creativity, naturalness, and melodiousness 8. Studies have also employed pairwise comparisons, where human participants listen to two music clips and indicate their preference based on overall quality or specific attributes like text-audio alignment 1. Furthermore, the visual analysis of musical scores, chord progressions, and spectrograms by music experts is considered a valuable subjective evaluation method, providing insights into the structural and harmonic aspects of the generated music 8.

While subjective evaluations are indispensable due to the limitations of objective metrics in fully capturing human perception, they are not without their challenges. Conducting large-scale subjective user studies can be resource-intensive, requiring significant time and effort in recruiting participants, designing experiments, and analyzing the collected data 6. Moreover, achieving high levels of reproducibility can be difficult due to variations in listener backgrounds, preferences, and the specific experimental conditions 6. Therefore, careful consideration of experimental design and the characteristics of the user group is paramount to ensure the reliability and validity of subjective evaluation outcomes 8. The practical challenges associated with subjective evaluations highlight the need for exploring more efficient and potentially automated methods to complement these human-centric approaches.

The criteria employed in subjective evaluations of AI-generated music are diverse and often tailored to the specific goals of the research. For instance, evaluations may focus on musicality, novelty, and the emotional impact of the generated pieces 8. Specific parameters such as creativity, naturalness, and melodiousness are frequently assessed using Likert scales 8. In the context of the AI Music Generation Challenge 2021, expert judges evaluated submissions based on dance ability, stylistic coherence, formal coherence, and playability 9. Similarly, the 2022 challenge utilized criteria like structure and melody, assessed by human experts familiar with the specific genre of Irish traditional music 10. The MusicEval dataset, designed for text-to-music evaluation, employs "overall musical impression" and "alignment with the text prompt" as key dimensions on which expert raters provide their assessments 11. This variety in evaluation criteria underscores the multifaceted nature of music quality and the diverse aspects that researchers aim to capture through subjective assessments.

**3. Objective Metrics in AI Music Assessment**

In addition to subjective evaluations, objective metrics offer computational approaches to assess the quality of AI-generated music by analyzing its various characteristics and generating quantitative measures. These metrics can be broadly categorized into model-based metrics, which evaluate the performance of the AI model itself (e.g., training loss, accuracy), and music domain metrics (MDM), which analyze the musical properties of the generated output 6. MDMs can be further classified into pitch-related metrics (e.g., pitch entropy, tonal stability), rhythm-related metrics (e.g., rhythmic complexity, tempo consistency), harmony-related metrics (e.g., chord progression analysis, harmonic consonance), and style-related metrics (e.g., genre classification accuracy) 6. Researchers have also developed innovative musicological objective assessment metrics that have demonstrated reproducibility in their application 8. A prominent objective metric used to quantify the perceptual quality of generated audio is the Frechet Audio Distance (FAD) 1. FAD measures the distributional similarity between a set of generated audio and a set of ground-truth audio by computing the Frechet distance between their audio embeddings. Various audio embedding models, such as VGGish, PANN, CLAP, and EnCodec, can be used to compute these embeddings, resulting in different variants of the FAD metric (e.g., FAD-VGG, FAD-CLAP) 1. For tasks involving text-to-music generation, cosine similarity between the text prompt and the generated audio embeddings, often derived from models like MS-CLAP and LAION-CLAP, is used to assess the alignment between the textual description and the musical output 1. Additionally, the Inception Score (IS) is another objective metric sometimes employed in evaluating generative models, including those for music 7.

Despite the scientific rigor offered by objective metrics, a significant challenge lies in their correlation with subjective human perception 6. Many objective metrics lack direct interpretability in terms of how they relate to the musical qualities that humans value 6. For instance, a high score on a particular objective metric might not necessarily translate to a piece of music that is perceived as high quality by human listeners. However, recent research has shown that certain objective metrics exhibit a stronger correlation with human preferences than others 13. A study benchmarking various music generation models and objective metrics against human preferences found that CLAP models trained on music data demonstrated the highest correlation with human evaluations for both overall music quality and text-audio alignment 1. This suggests that carefully selected and trained objective metrics can serve as valuable proxies for human judgment in specific contexts. Furthermore, a comprehensive study aimed at ranking state-of-the-art music generation models and metrics based on human preference represents a significant step towards identifying more reliable objective evaluation methods 1.

A persistent issue in the application of objective metrics is the difficulty in interpreting their scores in relation to the subjective experience of music 6. While these metrics can provide quantitative data on various musical features, understanding what a specific score signifies in terms of perceived musicality, emotional impact, or aesthetic appeal remains a considerable hurdle. Bridging this gap between the computational analysis of objective metrics and the qualitative experience of human listeners is crucial for the effective use of these tools in guiding the development and evaluation of AI music generation models.

**4. The Challenge of Evaluating Creativity and Aesthetics**

Evaluating creativity in AI-generated music presents a complex challenge that encompasses both philosophical and practical considerations. Creativity is often defined by criteria such as novelty, originality, and value 8. Novelty refers to the uniqueness of a musical piece, originality involves demonstrating an understanding of musical traditions while introducing something new, and value encompasses the emotional, cultural, aesthetic, or intellectual resonance of the music 8. Current evaluation methods may struggle to fully capture these multifaceted aspects of creativity, and the definition of creativity itself can vary across different audiences, making effective assessment particularly difficult 8. Research suggests that certain quantifiable musical features, such as the length of melodies, the number of distinct pitches used, and the information content of pitch intervals, can predict creativity ratings to some extent 14. Furthermore, the concept of divergent thinking, which involves exploring multiple possible solutions, is relevant to musical creativity, and advancements in generative AI are being leveraged to automate the scoring of tests designed to measure divergent thinking 14. The role of AI in expanding artistic boundaries and redefining authorship is also a subject of ongoing discussion 16. Despite these efforts, creating objective tools that accurately assess the creative merit of AI-generated music remains a significant challenge, often necessitating reliance on subjective evaluations 6.

The evaluation of aesthetics in AI music is inherently linked to subjective appreciation. Aesthetic judgments are deeply personal and are influenced by individual experiences, cultural backgrounds, and emotional responses to music 6. Assessments of musicality, emotional impact, and overall preference in listening tests directly relate to the aesthetic qualities of the generated music 8. The subjective evaluative nature of music means that what one listener finds aesthetically pleasing, another may not 6. This inherent subjectivity makes it challenging to standardize aesthetic evaluation or to fully capture it through objective metrics.

Given the complexities of evaluating creativity and aesthetics, the role of human expertise remains crucial. Music experts can provide valuable insights through visual analysis of musical elements and by applying their deep understanding of music theory and history to assess the novelty and originality of AI-generated compositions 8. The AI Music Generation Challenges, for example, rely heavily on expert human judges to evaluate various aspects of generated music, including its creativity and musicality within specific traditional music genres 9. Similarly, the MusicEval dataset specifically utilizes ratings from music experts to evaluate the overall musical impression and the alignment of the music with textual prompts 11. These experts possess the nuanced understanding necessary to discern the subtle qualities that contribute to the creative and aesthetic value of music 19.

**5. Assessing Out-of-Distribution AI Music**

Evaluating the generalization capabilities of AI music models, particularly their performance on data that differs significantly from their training distribution (out-of-distribution or OOD data), is essential for assessing their robustness and real-world applicability. Current evaluation methods often face limitations in this regard, particularly when it comes to generalizing across different music genres and cultural styles 8. A significant concern is the underrepresentation of music from the Global South in the datasets used to train AI music models. Research indicates that approximately 86% of the total dataset hours used in AI music generation research focus on music from the Global North, while genres from the Global South account for only 14.6% 20. This data imbalance raises concerns about cultural homogenization and the potential reinforcement of Western cultural dominance in AI-generated music 20.

Specific methodologies for evaluating out-of-distribution AI music within the music domain appear to be an area requiring further research. One potential approach involves the development of genre-specific evaluation metrics, as a model performing well in one genre might not necessarily do so in another 8. The general concept of out-of-distribution detection, such as the "Pseudo Outlier Exposure" method explored in other AI domains, might offer some insights for adaptation to music evaluation 21. However, there is a clear need for more dedicated research into specific metrics and methodologies tailored to assess the quality and appropriateness of AI-generated music in OOD scenarios.

Evaluating OOD music, especially across different cultural contexts, necessitates sensitivity to cultural nuances and a deep understanding of what constitutes authentic or appropriate style transfer. Simply generating music that sounds different from the training data is insufficient; the generated music must be stylistically coherent and culturally relevant within the target domain 20. The significant underrepresentation of diverse musical styles in training data can lead to models that struggle to generate culturally authentic music outside of their primary training distribution. Therefore, evaluation in this context should consider not only technical musical features but also the cultural appropriateness and potential for cultural homogenization 20.

**6. Benchmarks and Leaderboards in AI Music Generation**

Benchmarks and leaderboards serve as valuable resources for comparing the performance of different AI models across various tasks. In the domain of AI music generation, while there are leaderboards and benchmarks for related audio tasks such as Text-to-Speech (TTS Arena) and Automatic Speech Recognition (Open ASR Leaderboard) 22, as well as general categories like "Generative AI Leaderboards" and "Creativity and Generative AI Benchmarks" that could potentially include music generation 23, there is a relative scarcity of established, widely recognized benchmarks specifically dedicated to comprehensive quality evaluation in AI music generation. Some benchmarks exist for specific audio tasks like sound demixing 24. Additionally, research studies that benchmark music generation models and objective metrics against human preferences, such as the one discussed earlier 1, effectively create a ranking that can serve as a benchmark within the scope of the study.

A significant factor contributing to the limited number of standardized benchmarks in AI music generation is the lack of standardization in evaluation metrics and datasets 7. The absence of common reference points makes it challenging to directly compare the performance of different AI music generation models across various research efforts. This lack of standardization hinders the progress of the field by making it difficult to track advancements and identify the most effective approaches.

The AI Music Generation Challenge series represents a notable initiative in the field. Running annually since 2020, these challenges focus on evaluating AI music generated in specific traditional music styles through the assessment of expert human judges 9. Focusing on styles such as Irish double jigs (2020), Swedish slängpolska (2021), Irish reels (2022), and even the generation of artificial music traditions (2023), these challenges provide a valuable framework for evaluation within their specific domains and foster discussions about the application of AI in music.

The MusicEval dataset is another significant contribution towards standardized evaluation, specifically in the text-to-music domain 7. This dataset comprises a collection of music clips generated by 31 different models in response to 384 text prompts, along with over 13,000 ratings from 14 music experts. The experts evaluated the music based on overall musical impression and alignment with the text prompts. MusicEval serves as a valuable resource for training and benchmarking automatic evaluation models for text-to-music systems.

**Table 1: Overview of AI Music Generation Challenges**

| **Year** | **Focus** | **Key Evaluation Criteria** | **Use of Human Experts** |
| --- | --- | --- | --- |
| 2020 | Irish Double Jigs | Plagiarism, rhythm, pitch range, mode/accidentals (rejection); melody, structure, playability, memorableness, interestingness (rating) | Yes |
| 2021 | Swedish Slängpolska | Plagiarism, meter, rhythm (rejection); danceability, stylistic coherence, formal coherence, playability (rating) | Yes |
| 2022 | Irish Reels | Plagiarism, rhythm, mode/accidentals (rejection); structure, melody (rating); tune titling | Yes |
| 2023 | Artificial Music Traditions | Documentary evidence, technical process, reflection on cultural appropriation, data usage, intentions | Yes (curators) |

**Table 2: Summary of the MusicEval Dataset**

| **Metric** | **Value** |
| --- | --- |
| Number of Music Clips | 2,748 |
| Number of Text Prompts | 384 |
| Number of Models | 31 |
| Number of Expert Ratings | 13,740 |
| Evaluation Dimension 1 | Overall Musical Impression |
| Evaluation Dimension 2 | Textual Alignment |

**7. Comprehensiveness of Current Evaluation Frameworks**

Current evaluation frameworks for AI-generated music are increasingly recognizing the need for a balanced approach that integrates both human-centric and computational methods to effectively assess the multifaceted nature of music quality 6. While objective metrics provide valuable quantitative analysis of musical features, they often fall short in capturing the subjective aspects of music appreciation 7. Conversely, subjective evaluations, although crucial for understanding human perception, can be limited by scalability and reproducibility issues 7. The AI Music Generation Challenges demonstrate a strong reliance on expert human judgment to evaluate various aspects of generated music within specific traditional genres 9. The MusicEval dataset, with its expert ratings on a diverse set of models and prompts, represents a significant step towards creating high-quality, human-annotated data for evaluating text-to-music systems 7.

The coverage of different aspects of music quality varies across current evaluation methods. Objective metrics typically focus on technical and structural elements such as pitch, rhythm, and harmony 6. Subjective evaluations, on the other hand, can encompass more abstract qualities like emotional impact and aesthetic appeal 8. The AI Music Generation Challenges often evaluate aspects like danceability and stylistic coherence, reflecting a more holistic view of music quality within the context of traditional music 9. MusicEval focuses on overall musical impression and textual alignment, addressing both the perceptual quality of the music and its semantic coherence with the given text 11. While progress has been made in evaluating various dimensions of music quality, achieving comprehensive coverage that adequately captures all relevant aspects remains an ongoing challenge.

The consideration of user context and the intended purpose of the generated music is an area that could be more explicitly incorporated into future evaluation frameworks. While subjective evaluations inherently involve listener preferences, the specific context in which the music will be used (e.g., background music for a video, a standalone musical piece for listening pleasure) and the intended audience can significantly influence quality judgments. Future research might benefit from developing evaluation protocols that take these factors into account to provide a more nuanced assessment of AI-generated music.

**8. Key Findings from Recent Surveys and Research**

Recent research on the evaluation of AI-generated music reveals several key findings. Subjective evaluation, relying on human listeners, remains a vital component due to the inherent subjectivity of music, but it faces challenges related to the time and resources required, as well as the difficulty in achieving high reproducibility. Objective metrics offer computational analysis but often struggle to correlate strongly with human perception of music quality. Evaluating creativity and aesthetics in AI music is particularly complex, often requiring the nuanced judgment of human experts. Assessing the ability of AI models to generate high-quality music outside of their training data (out-of-distribution) is a critical area that needs further development, especially considering the potential for cultural biases in training datasets. While benchmarks and leaderboards exist for related audio tasks and general generative AI, standardized resources specifically for AI music generation are still in the early stages of development. However, initiatives like the AI Music Generation Challenge and the creation of datasets like MusicEval are significant steps towards addressing this gap. Current evaluation frameworks are evolving towards a more integrated approach, combining the strengths of both subjective and objective methodologies. Achieving a comprehensive evaluation that covers all relevant aspects of music quality and considers the context of use remains an ongoing challenge in the field.

Emerging trends in AI music evaluation include a growing emphasis on creating datasets with detailed human annotations, reflecting the importance of subjective perception. There is also an increasing focus on improving the correlation between objective metrics and human preferences to develop more reliable computational assessment tools. The AI Music Generation Challenges continue to serve as important platforms for evaluating creative AI in specific musical domains, fostering community engagement and driving research in this area. Furthermore, researchers are actively exploring new objective metrics that aim to better capture the perceptual qualities of music, moving beyond traditional feature-based approaches. The field is demonstrating a clear direction towards developing more reliable, valid, and comprehensive methods for evaluating AI-generated music, with a concerted effort to bridge the gap between computational analysis and the rich, subjective experience of music.

**9. Gaps and Limitations in AI Music Evaluation**

Current approaches to evaluating AI-generated music still exhibit several gaps and limitations, particularly concerning the assessment of creativity and aesthetics. Defining and quantifying these inherently subjective qualities through objective metrics remains a significant hurdle. While human evaluations can provide valuable insights, they are susceptible to biases and may not always be consistent across different listeners or contexts. There is a lack of universally accepted metrics that can fully capture the novelty, originality, and emotional resonance that define musical creativity and aesthetic appeal. Accurately and reliably evaluating the artistic merit and emotional impact of AI-generated music thus remains a substantial open challenge.

Evaluating the performance of AI music models on out-of-distribution data also presents considerable challenges. The lack of diverse datasets that adequately represent a wide range of musical styles and cultural traditions limits our ability to assess how well these models can generalize. Furthermore, there is a need for established metrics specifically designed to assess the cultural appropriateness and stylistic authenticity of AI-generated music in OOD scenarios. Determining whether a model has truly learned to generalize or is simply memorizing patterns from its training data is also a significant difficulty in this area. The evaluation of AI music's ability to generalize beyond its training data, especially to less represented musical styles and cultural contexts, requires more focused attention and research.

Beyond these specific areas, general gaps and limitations persist in the current approaches to AI music evaluation. A lack of standardization in both metrics and evaluation protocols makes it difficult to directly compare the performance of different models across different studies. Achieving high levels of reproducibility in subjective evaluations remains a challenge due to the inherent variability in human perception. Many existing objective metrics suffer from limited interpretability, making it hard to understand their scores in relation to perceived musical quality. Finally, there is a need for more comprehensive evaluation frameworks that explicitly consider the user context and the intended purpose of the generated music, as these factors can significantly influence how the music is perceived and evaluated. The field still lacks a unified and comprehensive framework that addresses both the technical and artistic aspects of AI-generated music across various contexts.

**10. Conclusion and Future Directions: Towards a More Holistic and Standardized Approach to Evaluating AI-Generated Music**

This report has highlighted the significant advancements in AI-generated music and the crucial role of robust evaluation methods in guiding the progress of this field. While both subjective and objective metrics have contributed to our understanding of AI music quality, several limitations and gaps remain. Subjective evaluations, though essential for capturing human perception, are resource-intensive and lack full reproducibility. Objective metrics offer quantifiable analysis but often struggle to correlate with human aesthetic judgments. Evaluating complex qualities like creativity and aesthetics, as well as the generalization capabilities of AI models to out-of-distribution data, presents ongoing challenges. The lack of standardized benchmarks and evaluation protocols further complicates the comparison of different AI music generation systems.

To address these limitations and move towards a more holistic and standardized approach to evaluating AI-generated music, several directions for future research are recommended. There is a need to develop more robust and interpretable objective metrics that exhibit a stronger correlation with human perceptual experiences of music. Creating larger and more diverse datasets for training and evaluation, with a particular focus on underrepresented musical traditions, is crucial for improving the generalization capabilities of AI models. Investigating novel methods for evaluating creativity and aesthetics, potentially drawing insights from music theory, psychology, and cognitive science, should be a priority. The development of standardized benchmarks and evaluation protocols would greatly facilitate the comparison of different AI music generation models and track progress in the field. Future evaluation frameworks should also explicitly consider the user context, intended purpose, and cultural relevance of the generated music to provide more nuanced assessments. Fostering interdisciplinary collaboration between AI researchers, musicologists, musicians, and cognitive scientists will be essential for advancing the field of AI music evaluation. Finally, exploring automated or semi-automated methods for subjective evaluation could help improve scalability and reduce the resource demands associated with human-centric assessments. Continued research in these areas will be vital for the advancement and responsible integration of AI-generated music into the broader music industry.

#### Works cited

1. tik-db.ee.ethz.ch, accessed March 20, 2025, <https://tik-db.ee.ethz.ch/file/e2394c9aba4e90f07e4f1bd8eedc823e/>
2. AI-Generated Music Detection and its Challenges - arXiv, accessed March 20, 2025, <https://arxiv.org/html/2501.10111v1>
3. AI in Music Market Size, Share, Trend | CAGR of 25.8%, accessed March 20, 2025, <https://market.us/report/ai-in-music-market/>
4. AI in Music Industry Statistics 2025: Market Growth & Trends - Artsmart.ai, accessed March 20, 2025, <https://artsmart.ai/blog/ai-in-music-industry-statistics/>
5. Artificial intelligence in music: recent trends and challenges - ResearchGate, accessed March 20, 2025, <https://www.researchgate.net/publication/385881724_Artificial_intelligence_in_music_recent_trends_and_challenges>
6. A Comprehensive Survey for Evaluation Methodologies of AI-Generated Music, accessed March 20, 2025, <https://www.researchgate.net/publication/373271455_A_Comprehensive_Survey_for_Evaluation_Methodologies_of_AI-Generated_Music>
7. MusicEval: A Generative Music Dataset with Expert Ratings for Automatic Text-to-Music Evaluation † These authors contributed equally to this work. ∗ Corresponding author. This work has been supported by the National Key R&D Program of China (Grant No.2022ZD0116307) and NSF China (Grant No.62271270). - arXiv, accessed March 20, 2025, <https://arxiv.org/html/2501.10811v1>
8. [2308.13736] A Comprehensive Survey for Evaluation Methodologies of AI-Generated Music, accessed March 20, 2025, <https://arxiv.org/abs/2308.13736>
9. The Ai Music Generation Challenge 2021: Summary and Results - YouTube, accessed March 20, 2025, <https://www.youtube.com/watch?v=kTKGAOWYuzw>
10. The Ai Music Generation Challenge 2022: Summary and Results - AIMC 2023 - PubPub, accessed March 20, 2025, <https://aimc2023.pubpub.org/pub/ynxz9uu7>
11. [Literature Review] MusicEval: A Generative Music Corpus with Expert Ratings for Automatic Text-to-Music Evaluation - Moonlight, accessed March 20, 2025, <https://www.themoonlight.io/review/musiceval-a-generative-music-corpus-with-expert-ratings-for-automatic-text-to-music-evaluation>
12. [Revue de papier] MusicEval: A Generative Music Corpus with Expert Ratings for Automatic Text-to-Music Evaluation - Moonlight, accessed March 20, 2025, <https://www.themoonlight.io/fr/review/musiceval-a-generative-music-corpus-with-expert-ratings-for-automatic-text-to-music-evaluation>
13. Benchmarking Music Generation Models and Metrics via Human Preference Studies - OpenReview, accessed March 20, 2025, <https://openreview.net/pdf?id=105yqGIpVW>
14. Full article: Sequence Generation and Evaluation: A Novel Assessment of Musical Creativity, accessed March 20, 2025, <https://www.tandfonline.com/doi/full/10.1080/10400419.2024.2425494>
15. Assessing Creativity Across Multi-Step Intervention Using Generative AI Models, accessed March 20, 2025, <https://learning-analytics.info/index.php/JLA/article/view/8571>
16. The Emerging “AI Artists”: Breaking the Metacrisis and the Fear of Losing Human Creativity, accessed March 20, 2025, <https://www.sciepublish.com/article/pii/418>
17. Protecting Human Creativity in AI-Generated Music with the Introduction of an AI-Royalty Fund | GRUR International | Oxford Academic, accessed March 20, 2025, <https://academic.oup.com/grurint/article/73/12/1137/7832810>
18. A Comprehensive Survey for Evaluation Methodologies of AI-Generated Music | AI Supported Adaptive Sound Design | Zotero, accessed March 20, 2025, <https://www.zotero.org/groups/5567041/ai_supported_adaptive_sound_design/items/itemKey/DBTMHTK7>
19. Piano performance evaluation dataset with multilevel perceptual features - PMC, accessed March 20, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11450231/>
20. Missing Melodies: AI Music Generation and its ”Nearly” Complete Omission of the Global South - arXiv, accessed March 20, 2025, <https://arxiv.org/html/2412.04100>
21. ‪Jaeyoung Kim‬ - ‪Google 학술 검색‬, accessed March 20, 2025, <https://scholar.google.co.kr/citations?user=3j4Q6_4AAAAJ&hl=ko>
22. Leaderboards and benchmarks - a clefourrier Collection, accessed March 20, 2025, <https://huggingface.co/collections/clefourrier/leaderboards-and-benchmarks-64f99d2e11e92ca5568a7cce>
23. Exploring AI Benchmarks & Leaderboards | dasarpAI, accessed March 20, 2025, <https://dasarpai.com/dsblog/exploring-ai-benchmarks-and-leaderboards>
24. Benchmarks and leaderboards for sound demixing tasks - arXiv, accessed March 20, 2025, <https://arxiv.org/html/2305.07489v2>
25. [2305.07489] Benchmarks and leaderboards for sound demixing tasks - arXiv, accessed March 20, 2025, <https://arxiv.org/abs/2305.07489>
26. boblsturm/aimusicgenerationchallenge2022: The Ai Music Generation Challenge 2022 - GitHub, accessed March 20, 2025, <https://github.com/boblsturm/aimusicgenerationchallenge2022>
27. The Ai Music Generation Challenge 2020: Double Jigs in the Style of O'Neill's ``1001'', accessed March 20, 2025, <https://www.researchgate.net/publication/355509723_The_Ai_Music_Generation_Challenge_2020_Double_Jigs_in_the_Style_of_O'Neill's_1001''>
28. The Ai Music Generation Challenge 2020: Double Jigs in the Style of O'Neill's ``1001'', accessed March 20, 2025, <https://www.jcms.org.uk/article/id/950/#!>
29. The Ai Music Generation Challenge 2023: "Artificial Music Traditions" - THE POSTHUMANITIES HUB, accessed March 20, 2025, <https://posthumanitieshub.net/2023/03/15/the-ai-music-generation-challenge-2023-artificial-music-traditions/>
30. Udio | AI Music Generator - Official Website, accessed March 20, 2025, <https://www.udio.com/>
31. AI Music Creativity 2022, accessed March 20, 2025, <https://2022.aimusiccreativity.org/>
32. The Ai Music Generation Challenge 2023 - Google Groups, accessed March 20, 2025, <https://groups.google.com/a/tensorflow.org/g/magenta-discuss/c/clJUNrkT_RQ>
33. VirtuosoTune: Hierarchical Melody Language Model - IEIE SPC - IEIE Transactions on Smart Processing & Computing, accessed March 20, 2025, <http://ieiespc.org/ieiespc/XmlViewer/f424058>
34. The Ai Music Generation Challenge 2022 : Summary and Results - kth .diva, accessed March 20, 2025, [https://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A1795614](https://kth.diva-portal.org/smash/record.jsf?pid=diva2:1795614)
35. The Ai Music Generation Challenge 2021: Summary and Results - DiVA portal, accessed March 20, 2025, <https://www.diva-portal.org/smash/get/diva2:1753806/FULLTEXT01.pdf>
36. boblsturm/aimusicgenerationchallenge2021: The Ai Music Generation Challenge 2021, accessed March 20, 2025, <https://github.com/boblsturm/aimusicgenerationchallenge2021>
37. The AI Music Generation Challenge 2020: Summary and Results | Folk the Algorithms, accessed March 20, 2025, <https://highnoongmt.wordpress.com/2020/11/20/the-ai-music-generation-challenge-2020-summary-and-results/>
38. [2501.10811] MusicEval: A Generative Music Corpus with Expert Ratings for Automatic Text-to-Music Evaluation - arXiv, accessed March 20, 2025, <https://arxiv.org/abs/2501.10811>
39. MusicEval: A Generative Music Corpus with Expert Ratings for Automatic Text-to-Music Evaluation - ResearchGate, accessed March 20, 2025, <https://www.researchgate.net/publication/388232526_MusicEval_A_Generative_Music_Corpus_with_Expert_Ratings_for_Automatic_Text-to-Music_Evaluation?_tp=eyJjb250ZXh0Ijp7InBhZ2UiOiJzY2llbnRpZmljQ29udHJpYnV0aW9ucyIsInByZXZpb3VzUGFnZSI6bnVsbH19>
40. ‪Hui Wang (王卉)‬ - ‪Google Scholar‬, accessed March 20, 2025, <https://scholar.google.com/citations?user=8witCh8AAAAJ&hl=en>