

MUSGO: A COMMUNITY-DRIVEN FRAMEWORK FOR ASSESSING OPENNESS IN MUSIC-GENERATIVE AI

Roser Batlle-Roca¹ Laura Ibáñez-Martínez¹ Xavier Serra¹
Emilia Gómez^{1,2} Martín Rocamora¹

¹ Music Technology Group, Universitat Pompeu Fabra, Barcelona Spain

² Joint Research Centre, European Commission, Seville, Spain

{roser.batlle, laura.ibanez, martin.rocamora}@upf.edu

ABSTRACT

Since 2023, generative AI has rapidly advanced in the music domain. Despite significant technological advancements, music-generative models raise critical ethical challenges, including a lack of transparency and accountability, along with risks such as the replication of artists' works, which highlights the importance of fostering openness. With upcoming regulations such as the EU AI Act encouraging open models, many generative models are being released labelled as 'open'. However, the definition of an open model remains widely debated. In this article, we adapt a recently proposed evidence-based framework for assessing openness in LLMs to the music domain. Using feedback from a survey of 110 participants from the Music Information Retrieval (MIR) community, we refine the framework into MusGO (Music-Generative Open AI), which comprises 13 openness categories: 8 *essential* and 5 *desirable*. We evaluate 16 state-of-the-art generative models and provide an openness leaderboard that is fully open to public scrutiny and community contributions. Through this work, we aim to clarify the concept of openness in music-generative AI and promote its transparent and responsible development.

1. INTRODUCTION

Music-generative AI is introducing critical ethical concerns, particularly regarding its impact on creative processes and authorship, potential legal issues from data misuse, and disruptions to existing business and intellectual property (IP) models [1–5]. Furthermore, these technologies usually exhibit a Western cultural bias, undermining diversity in musical expression and homogenising music genres. Additionally, they often require substantial computational resources, contributing to energy consumption amid the climate crisis [4, 6]. A further concern is the lack of transparency in these models, which limits scrutiny of

their behaviour, attribution, and decision-making [5], as well as the potential replication of training data [7–9].

To address these challenges, we advocate for the use of open models. Open models have a vital role in open science, promoting transparency, accountability, and innovation [10–12]. Alongside academic efforts [13], tech companies are increasingly labelling AI systems as open. However, several initiatives provide only a few components, such as model weights and inference code [11], which limits reuse and transparency, and leads to criticism of 'openwashing'. Yet, defining openness for AI models is challenging and remains a subject of active debate. In fact, open model status is becoming highly attractive with the legal exceptions in the EU AI Act, making it timely to examine the requirements needed to attain it [11, 14].

The absence of a defined methodology for evaluating openness in music-generative models creates a potential gap in their assessment. Consequently, we pose the question: **how can openness be assessed in music-generative AI?** We draw upon and adapt the work by Liesenfeld and Dingemanse (2024) [11], which introduces an evidence-based method for evaluating openness in large language models (LLMs). Their proposed framework is composite and graded, consisting of multiple categories, each with three levels of openness. Building on and tailoring this approach for the music domain, we introduce **MusGO** (Music-Generative Open AI): a community-driven framework for assessing openness in music-generative AI. We develop MusGO by actively involving the Music Information Retrieval (MIR) community through a survey and refining the criteria based on their feedback.

Our main contributions include (1) a reproducible **framework**¹ for evaluating openness in music-generative AI, (2) a **leaderboard**² showcasing openness across 16 models, and (3) a fully open **repository**³ to enable public scrutiny, community contribution and future updates. With this, we aim to support future work by providing a more nuanced understanding of openness in music-generative AI, offering a fine-grained evidence-based approach to inform decisions and promote responsible practices.



© R. Batlle-Roca, L. Ibáñez-Martínez, X. Serra, E. Gómez, and M. Rocamora. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** R. Batlle-Roca, L. Ibáñez-Martínez, X. Serra, E. Gómez, and M. Rocamora, "MusGO: A Community-Driven Framework for Assessing Openness in Music-Generative AI", in *Proc. of the 26th Int. Society for Music Information Retrieval Conf.*, Daejeon, South Korea, 2025.

¹ https://roserbatlleroca.github.io/MusGO_framework/framework.html

² https://roserbatlleroca.github.io/MusGO_framework/index.html

³ https://github.com/roserbatlleroca/MusGO_framework

2. BACKGROUND AND RELATED WORK

2.1 Defining ‘open’ models

Documentation efforts in AI have supported model transparency by disclosing development processes, data sources, and model attributes [15–18]. However, defining openness in AI is challenging, as it involves multiple components (e.g., source code, documentation, model weights, training data) [13] and requires balancing openness ideals with practical concerns such as data privacy, security, and IP rights [19]. Recent initiatives have contributed to shaping such a definition. One of these is the Open Source Initiative (OSI), which has led a public consultation to adapt its widely accepted definition of open-source software to AI systems, resulting in the release of the Open Source AI Definition (OSAID) 1.0 in October 2024 [20]. It requires an AI system to provide access to its source code and model weights, along with sufficient information about the training data. The goal is to ensure these models can be freely used, modified, and shared, thereby promoting transparency and collaboration within the AI community.

While OSAID represents an important step toward setting a standard for AI openness, it has also sparked discussion and criticism [21–23]. Some argue that the stringent requirements may be challenging to meet, possibly excluding models that, although not fully compliant, still contribute meaningfully to the open-source ecosystem. Another concern raised by open data advocates is that OSAID does not require the training data itself to be shared, only detailed information about it, which is sometimes seen as a form of ‘open-washing’. This is particularly relevant in music, where much of the data, such as recordings, compositions, and lyrics, is owned by rights holders and often cannot be made publicly available. The freedom to use a model for any purpose also raises concerns about unintentional copyright violations or the deliberate imitation of an artist’s style without their consent [5, 6, 24]. In contrast, Responsible AI Licenses (RAIL) propose a more nuanced form of openness by including restrictions on certain uses.

Among other regulatory efforts, the European Union’s AI Act [25] has granted some exemptions to models released under a free and open-source license, as specified in Article 53 (1)(a) and (b) [26]. The public availability requirement appears to cover “(...) *its parameters, including the weights, the information on the model architecture, and the information on model usage*, with no explicit mention of the training data [27]. The *General-Purpose AI Code of Practice* is expected to provide further details on the obligations under Articles 53 and 55 for different ways of releasing general-purpose AI models, including open sourcing, although the document is still in draft form [28].

Rather than conceiving openness as a binary status, other proposals are tiered, with openness distributed across different dimensions [14, 29]. This aligns with prior work on AI accountability and transparency frameworks [30–33]. In this view, a model’s openness is determined by the relative openness of its individual components, accommodating cases where “not all criteria are fully attained.

2.2 How to assess openness?

Building on the idea that openness is not a binary attribute but rather a composite of various elements, each varying across different degrees, several initiatives have aimed to create structured, tiered, gradient-based frameworks to evaluate it [11, 34–36].

For instance, the Linux Foundation’s Model Openness Framework (MOF) [35] distinguishes three tiers of open AI systems. It promotes completeness and openness for reproducibility, transparency, and usability, categorising AI models based on their level of openness and ensuring the release of key artefacts like model architecture, training data, and evaluation results under open licenses. Similarly, the Foundation Model Transparency Index (FMTI) [34] evaluates transparency in foundation models using 100 indicators across three broad domains—upstream, model, and downstream—focusing on aspects such as data, compute resources, model capabilities, and downstream impact. While comprehensive, the FMTI does not allow individual data points to be scrutinised or contested, limiting the transparency and verifiability of its scores [11].

On the other hand, Eiras et al. [36, 37] have contributed valuable insights into the risks and opportunities associated with open-source generative AI models, proposing a taxonomy for evaluating their level of openness, particularly focusing on the availability of code and data. Although advocating for AI democratisation, they emphasise the need for openness to be balanced with responsible AI practices to avoid risks such as misuse, bias, and unintended consequences. Their evaluation framework highlights both technical openness (e.g., code and model weights availability) and ethical concerns (e.g., data privacy and model safety), offering a more comprehensive perspective on open-source AI compared to other approaches.

A particularly relevant framework is the one proposed by Liesenfeld and Dingemanse (2024) [11] which, building on previous work [29], offers an evidence-based method for evaluating openness in LLMs, and extends it to text-to-image models. The proposed framework is **composite**—it considers multiple dimensions when determining openness—and **graded**—meaning each category may have different levels of openness: *closed*, *partial* and *fully open*. Many of these dimensions align with those from other frameworks, particularly MOF. The approach has been tested since July 2023 in an open leaderboard⁴ that tracks the degree of openness of several LLMs. A distinguishing feature of this method is its community-driven approach, which allows for the inspection and appeal of the model’s assessment scores and their supporting evidence. However, while the framework is adaptable to various domains, it is essential to revise certain categories and include domain-specific elements to accurately assess openness [11]. In the case of music, unique challenges, such as the copyrighted nature of most training data, underscore the need for tailored criteria and domain-specific practices to ensure a comprehensive openness assessment.

⁴ <https://opening-up-chatgpt.github.io/>

3. OPENNESS FRAMEWORK FOR MUSIC AI

3.1 From LLMs to music

We adopted the evidence-based framework introduced by Liesenfeld and Dingemanse (2024) [11] and tailored it to the music domain. Our initial adaptation involved modifying references to LLMs to align with music-generative models, for example, by renaming LLM-oriented labels and excluding instruction tuning-related categories, which led to a revised framework of 11 categories from the original 14. To ensure the adapted framework reflected the perspectives and priorities of the MIR community, we aimed to gather feedback on each category’s intention and relevance, considering the risks of over-reliance on single features, such as access or licensing, to determine openness [11]. To this end, we conducted an anonymous online questionnaire within the MIR community.

3.2 MIR community survey

3.2.1 Methodology

The survey included a statement for each category reflecting the top open level (*fully open*). Participants were asked to assess the proposed statement by (1) **relevance**—how important they consider the category in determining the openness of a model—using a 5-point Likert scale (1 = *not relevant*, 5 = *very relevant*), and (2) **agreement**—whether they agreed, somewhat agreed, or disagreed with the proposed statement. They were encouraged to suggest modifications and refinements to the statements when they disagreed or had reservations about the criteria. The survey also included an open comments section for participants to propose new categories and raise unaddressed concerns.

The survey was open for 8 weeks and received 110 responses from participants linked to the MIR community. Although we intended broader representation, the sample was biased towards male academics in Europe and North America. However, this distribution aligns with the typical demographics of attendees at the International Society for Music Information Retrieval (ISMIR) conference.⁵

3.2.2 Results

Table 1 summarises the survey results, showing the median relevance score (M) and agreement levels for each category. Overall, participants mostly agreed with the proposed statements, indicating a general consensus on the proposed criteria. Categories 3: *Model Weights*, 4: *Code documentation* and 1: *Open Code* showed the highest levels of agreement (with over 78% selecting *Yes, I agree*), and were also rated highly relevant ($M_{1,3}=5$, $M_4=4$). For category 2: *Training data*, while relevance was high ($M_2=5$), there was significant discussion regarding the framing of the statement (30% selected *Yes, I somewhat agree*). These results suggest that elements related to code access and model reproducibility are of great interest to the community.

⁵ Complementary information on the survey, including a detailed analysis of the participants’ backgrounds and their responses, is available at: https://roserbatlloca.github.io/MusGO_framework/survey.html

Table 1: Results from the MIR community survey, showing for each category its median relevance and level of agreement (%) with the proposed statement. Note that the criteria correspond to the first adapted version (see 3.1).

Criteria	Relevance <i>Median</i>	Agreement (%)		
		<i>Yes, I agree.</i>	<i>Yes, I somewhat agree.</i>	<i>No, I do not agree.</i>
1: Open Code	5	78.18	20.00	1.82
2: Training Data	5	67.27	30.00	2.73
3: Model Weights	5	79.09	18.18	2.73
4: Code documentation	4	79.09	18.18	2.73
5: Architecture	4	62.73	30.00	7.27
6: Paper and Preprint	4	64.55	21.82	13.64
7: Model Card	4	54.55	30.00	15.45
8: Datasheet	4	64.55	28.18	7.27
9: Package	3	46.36	32.73	20.91
10: API	3	39.09	30.00	30.91
11: Licensing	4	62.73	28.18	9.09

In contrast, categories such as 9: *Package* and 10: *API*, showed lower and more varied agreement rates (46.36% and 39.09%, respectively) and lower relevance ($M_{9,10}=3$). This was reflected in the comments, which highlighted concerns about the additional effort required to provide these features and questioned their necessity for determining openness.

The survey also highlighted some categories with moderate agreement, such as 6: *Paper and Preprint*, 7: *Model Card* and 8: *Datasheet*, all with a median relevance score of 4, but with notable disagreement rates. This diversity of responses underscores the varying priorities among community members regarding detailed documentation across different areas of the model. Comments on these categories were key to understanding the community’s perspectives and refining the initial criteria.

3.3 Refining criteria based on community feedback

Participants’ feedback revealed gaps and disagreements that we addressed by carefully reviewing all comments, identifying recurring issues, and refining the criteria accordingly. The most substantial changes involved clarifying the descriptions of each category and defining what constituted *partial* and *fully open* for each case. We integrated music-specific nuances raised in the feedback, including the challenges of sharing IP-protected training data, the need to provide sonified examples, and the user experience needs of musicians, who may have limited technical skills and domain-specific software preferences, extending beyond the use of APIs. In addition, we held internal discussions with several researchers from the Music Technology Group (MTG), which provided further insights that helped refine the framework.

Following the survey results regarding category relevance, we introduced two distinct levels within our framework: **essential**, for categories considered key to assessing openness, and **desirable**, for those that enhance openness and provide additional value but are not indispensable. Furthermore, feedback highlighted the absence of a category addressing model evaluation procedures. In response, we introduced a new category, *Evaluation pro-*

cedure, covering evaluation data, metrics, and results on model performance. We also added the category *Supplementary material page* to acknowledge the importance of dedicated websites that provide usage instructions, model demonstrations and sonified outputs, which are essential in music research, where audio cannot be embedded within the paper itself. Additionally, we renamed some categories to better reflect their scope: *Open code* → *Source code*, *Architecture* → *Training procedure*, *Paper* → *Research Paper*, and *API* → *User-oriented application*. The renamed category *User-oriented application* also includes other user-experience approaches, such as real-time applications, recognising the need for accessible tools for musicians and composers. These music-specific adaptations were critical to make the framework relevant for music-generative models, which differ significantly from LLMs.

3.4 The MusGO framework

As a result of this modification and iteration process, we built the Music-Generative Open AI framework. MusGO comprises 13 categories: 8 essential and 5 desirable. The essential categories include: (1) *Source code*, (2) *Training data*, (3) *Model weights*, (4) *Code documentation*, (5) *Training procedure*, (6) *Evaluation procedure*, (7) *Research paper*, and (8) *Licensing*, while desirable categories are defined by (9) *Model card*, (10) *Datasheet*, (11) *Package*, (12) *User-oriented application*, and (13) *Supplementary material page*. Essential categories follow an openness-graded scale of 3 levels: closed (✗), partial (~) and fully (open) (✓). Instead, desirable categories are binary: whether that element exists (★) or not, as they are considered complementary add-ons to the model, rather than core components, which allows supporting any efforts made in these areas. A complete description of the criteria is available on our complementary website.

4. ASSESSING OPENNESS

4.1 Model selection

We selected 16 state-of-the-art music generation models:⁶ GANSynth [38], Jukebox [39], RAVE [40], Musika [41], Moûsai [42], MusicGen [43], MusicLM [44], VampNet [45], MusicLDM [46], Music ControlNet [47], Noise2Music [48], MeLoDy [49], DITTO-2 [50], Diff-A-Riff [51], JASCO [52], and Stable Audio Open [53]. Our selection encompasses a diverse set of music-generative models with waveform outputs, mainly focusing on those that explicitly claim to be ‘open’ or are associated with openness, alongside popular models that exhibit openness behaviours (e.g., publishing model weights, releasing research papers, or providing user-oriented applications). In addition to this, we aimed to cover a variety of architectures commonly used within the field, including Variational Autoencoders (VAEs) [54], Generative Adversarial Networks (GANs) [55], Transformers [56], Diffusion models [57], and novel methodologies such as Flow Matching [58].

⁶ Introduced from oldest to newest research paper release.

4.2 MusGO into practice

We evaluated the selected models using a structured methodology based on a checklist of fully open level requirements for each category. Each model was examined by one of the authors, including justification statements. Two other authors then reviewed the evaluations to ensure consistency and minimise bias. In case of discrepancy, we engaged in discussions and adjusted the assessment iteratively, following consensual qualitative research principles [59]. We gathered information from official publishers and maintainers and, when available, complemented it with third-party sources such as HuggingFace.⁷

4.3 Openness assessment

Figure 1 presents the openness evidence-based assessment of the 16 selected models, forming the basis of our openness leaderboard. Models are ordered using a weighted openness score (O), based on essential categories (E) and normalised to a 100-point scale. Considering survey findings, the three most relevant categories (E_1 , E_2 and E_3 , all with $M_{1,2,3}=5$) are weighted twice as much as the others. Note that the score is used for ordering purposes only, and we do not intend to reduce openness to a single value. When models achieve the same score, the order is determined by the highest number of fulfilled desirable categories.

The assessment reveals significant variation in openness across the considered models. The category *Training procedure* is the most open, with 11 models fully open and 5 partially open. Instead, the category *Training data* is the most closed, with only one model (Stable Audio Open [53]) achieving fully open status. Regarding source code, eight models provide full access, and two provide partial access. The remaining six models, besides not providing source code, also do not disclose model weights, code documentation, or apply appropriate licensing. Models classified as fully open for category *Model weights* also tend to provide comprehensive documentation of their code and are typically licensed under an OSI or RAIL license, reflecting a correlation between source code, model weights, code documentation, and licensing. Regarding the category *Evaluation procedure*, most models (11/16) are partially open, often lacking evaluation data or implementation details, while only five are fully open. Although all models analysed provide a research paper or equivalent technical report, only 11 meet the peer-reviewed and accessibility requirements. Desirable categories also show a substantial variability across models. The category *Datasheet* is the least fulfilled, with only one model (MusicLDM [46]) providing detailed and structured data documentation. Instead, all models include a supplementary material page, which demonstrates that accompanying resources with model demonstrations and sound examples has become a community norm. These results are made publicly available through a continuously updated leaderboard, helping to identify incomplete openness claims and potential cases of ‘open-washing’.

⁷ <https://huggingface.co/>

Project	Essential								Desirable				
	Source code	Training data	Model weights	Code documentation	Training procedure	Evaluation procedure	Research paper	License	Model card	Datasheet	Package	UX application	Supplementary material page
MusicGen Meta AI	✓	~	✓	✓	✓	✓	✓	✓	★		★	★	★
Stable Audio Open Stability AI	✓	✓	✓	✓	✓	✓	~	~	★		★		★
JASCO The Hebrew University of ...	✓	~	✓	✓	~	✓	✓	✓	★		★	★	★
GANSynth Google Magenta	✓	~	✓	✓	✓	~	✓	✓			★		★
Musika Johannes Kepler Universi...	✓	~	✓	✓	✓	~	✓	✓				★	★
MusicLDM University of California S...	~	~	✓	✓	✓	✓	✓	~	★	★			★
VampNet Northwestern University a...	✓	✗	✓	✓	✓	~	✓	✓				★	★
Jukebox OpenAI	✓	~	✓	✓	✓	~	~	~					★
RAVE IRCAM	✓	✗	✓	✓	✓	✓	~	~	★		★	★	★
Moūsai ETH Zürich, IIT Kharagpu...	~	~	✗	~	✓	~	✓	✓			★		★
Diff-A-Riff Sony Computer Science ...	✗	✗	✗	✗	✓	~	✓	✗					★
Music ControlNet Carnegie Mellon Universit...	✗	✗	✗	✗	✓	~	✓	✗					★
MusicLM Google Research and IR...	✗	~	✗	✗	~	~	~	✗					★
Noise2Music Google Research	✗	~	✗	✗	~	~	~	✗					★
DITTO-2 University of California S...	✗	✗	✗	✗	~	~	✓	✗					★
MeLoDy ByteDance	✗	✗	✗	✗	~	~	✓	✗					★

Figure 1: Openness leaderboard of 16 music-generative models assessed using the MusGO framework.

5. DISCUSSION

5.1 Music-generative open AI

Our assessment reveals a significantly diverse landscape of music-generative models in terms of openness, highlighting both notable efforts and significant room for improvement. This situation underscores different levels of commitment across the community, particularly regarding critical categories such as training data, source code, and licensing. Although the analysed models generally perform well in providing documentation, the level of detail is not always sufficient to guarantee complete transparency and accountability. Furthermore, desirable categories are less commonly addressed, especially datasheets, pointing to a gap in the information available about these models.

Our evaluation validates the MusGO framework as a comprehensive evidence-based resource for assessing openness in music-generative AI. Using a composite and graded framework allows for a more flexible and fine-grained perspective on model openness. For instance, while *Training procedure* is often the most open category, *Training data* remains critical, with the majority of models

lacking sufficiently detailed descriptions of the data used. However, we acknowledge that sharing training data in the music domain may pose significant challenges due to IP rights, which can lead to ethical and legal concerns when publishing this information. We address this issue by considering *Training data* fully open when direct access is restricted due to legal concerns, provided that detailed information about all sources is disclosed.

We present a continuously updated leaderboard showcasing openness assessments using the MusGO framework. It reflects the current state of openness in music-generative models and acts as a live collaborative platform for the community. Researchers, developers, and stakeholders are invited to actively contribute by evaluating new models and engaging in discussion. This approach enables tracking model evolution and integrating emerging aspects like controllability and real-time use. The repository provides guidelines on how to contribute, including the detailed framework criteria, instructions for issues and pull requests, and information on submissions review.⁸

⁸ https://github.com/roserbatlleroca/MusGO_framework/tree/main/projects/README.md

5.2 Beyond openness

While openness is crucial for ensuring transparency and accountability in music-generative models, it alone does not guarantee ethical behaviour or address broader implications. Several survey participants also raised this concern. Barnett [4] demonstrates that most literature on generative AI in music tends to overlook the potential negative ethical consequences of these models. However, our analysis suggests a shift with increasing attention focused on these concerns. Among the analysed models, 62.5% (10/16) include ethical considerations, with a notable rise in papers published after 2023 (83.3% of these models).

Openness can help reveal where ethical risks arise, especially in sensitive areas such as copyright. While sharing code and data supports transparency, it does not address the legality of the data used for training. For instance, MusicGen [43] and JASCO [52] obtained data through proper legal agreements, whereas Stable Audio Open [53] is the only model using fully available and accessible data. This highlights the need to balance openness with careful copyright considerations. Thus, MusGO helps identify and evaluate whether ethical claims are substantiated.

Similarly, openness can point out whether efforts to mitigate harmful or inappropriate uses have been implemented. Models like MusicLDM [46] and Stable Audio Open [53] try to mitigate risks by performing memorisation analysis to help ensure that training data is not inadvertently reproduced by their models. In addition, MusicLM [44] and Noise2Music [48] address the risks of plagiarism and cultural appropriation by incorporating safeguards to mitigate such issues. While MusGO does not directly evaluate these aspects, future versions of the framework could help emphasise risk mitigation strategies.

Another key concern is the representation and diversity of the training data. Open datasets may still be biased, with limited cultural diversity. For example, JASCO [52] acknowledges that its dataset is heavily Western-centric, which limits the diversity of the generated music. Similarly, Noise2Music [48] points out that biases in training data can misrepresent musical genres. These examples show that MusGO can help expose limitations in cultural diversity and representation by making dataset composition and documentation more accessible for scrutiny.

Finally, the economic impact of generative AI on musicians and creative professionals is also a critical concern. Open models increase accessibility, but they can potentially displace human creators. Many authors stress that generative models should enhance, not replace, human creativity. Yet, concerns remain about the disruption of traditional industries and the economic impact on musicians. This emphasises the need for openness to be paired with a broader discussion about the ethical, social, and economic implications of AI, including potential usage restrictions. Thus, ethical guidelines and safeguards are essential to ensure that AI-generated music does not perpetuate harmful stereotypes or infringe on the rights of original creators. MusGO offers a structured lens to review models' governance and serve as a foundation for ethical scrutiny.

5.3 Limitations

Evaluating the category *Training data* has proven particularly challenging. In the music domain, data accessibility is often restricted by IP rights and legal constraints, and even revealing training data might require a closed audit process to comply with these requirements. Incorporating such a level of detail into our framework would require a more complex approach, which may not be feasible in this context. Moreover, while hardware requirements are considered within the *Training procedure* category, their impact on openness has not been thoroughly analysed. Some models can be trained on personal computers, while others require extensive computational resources, complicating reproducibility and limiting accessibility. For models with lower resource demands, categories like *Training data* may become less relevant, shifting the focus toward enabling artists to train models on their own music.

In developing the MusGO framework, we surveyed the MIR community. However, we acknowledge that our sample was biased towards male academics based in Europe and North America. Expanding the sample in the future to include artists and other stakeholders, as well as perspectives from a broader range of regions, could provide valuable insights and lead to adaptations that would make the framework more inclusive.

While MusGO provides a robust assessment of openness, the leaderboard does not reflect the broader ethical and societal implications of generative AI. This remains a key direction for future improvement, especially as implications related to hardware requirements, real-time use, and accessibility are highly relevant to artists.

6. CONCLUSION

With the rise of music-generative AI, debates around the ethical implications of these models have intensified. Given the lack of transparency and accountability in these systems, we advocate for open models. Yet, what constitutes an open model remains undefined for music-generative AI. In this work, we adapt an existing openness framework for LLMs to music-generative AI. A community-specific survey helped us identify gaps and particular considerations unique for the music domain. As a result of tailoring the criteria based on the gathered feedback, we introduce the MusGO framework. We put MusGO into practice by analysing 16 state-of-the-art models, providing a public and updatable openness leaderboard. Our analysis reveals that openness in music-generative AI is a field in progress, where significant gaps remain in the availability of training data, model weights, and licensing. MusGO allows for a structured and flexible perspective on model openness by considering complementary characteristics. It reports on a model's status and highlights potential cases of 'open-washing'. As generative models continue to evolve, we strive for more consistent and thorough openness practices, particularly when sharing training data and source code, to foster transparency, accountability, and responsible development in music-generative AI.

7. ETHICS STATEMENT

This study involved a voluntary, anonymous online survey aimed at gathering feedback from the MIR community on a preliminary adapted openness framework. Participants were informed about the scope and purpose of the study, as well as the intended use of the collected data. Regarding study design, participant information and data protection, we adhered to the guidelines and recommendations of the Institutional Committee for Ethical Review of Projects (CIREP) at Universitat Pompeu Fabra. In line with the General Data Protection Regulation (GDPR) 2016/679 (EU), all responses were anonymised and stored securely. Participants were informed of their data rights, including access, rectification, deletion, and withdrawal of consent, in accordance with GDPR protocols.

8. ACKNOWLEDGMENTS

This work has been supported by *IA y Música: Cátedra en Inteligencia Artificial y Música* (TSI-100929-2023-1), funded by the Secretaría de Estado de Digitalización e Inteligencia Artificial and the European Union-Next Generation EU, and *IMPA: Multimodal AI for Audio Processing* (PID2023-152250OB-I00), funded by the Ministry of Science, Innovation and Universities of the Spanish Government, the Agencia Estatal de Investigación (AEI) and co-financed by the European Union. We thank our colleagues at the Music Technology Group at Universitat Pompeu Fabra for their thoughtful insights, constructive discussions and active engagement throughout the development of this work.

9. REFERENCES

- [1] F. Carnovalini and A. Rodà, “Computational creativity and music generation systems: An introduction to the state of the art,” *Frontiers in Artificial Intelligence*, vol. 3, 2020.
- [2] E. Gómez, M. Blaauw, J. Bonada, P. Chandna, and H. Cuesta, “Deep learning for singing processing: Achievements, challenges and impact on singers and listeners,” *arXiv preprint arXiv:1807.03046*, 7 2018. [Online]. Available: <https://arxiv.org/abs/1807.03046v1>
- [3] B. L. T. Sturm, M. Iglesias, O. Ben-Tal, M. Miron, and E. Gómez, “Artificial intelligence and music: Open questions of copyright law and engineering praxis,” *Arts*, vol. 8, p. 115, 9 2019.
- [4] J. Barnett, “The ethical implications of generative audio models: A systematic literature review,” *AIES 2023 - Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 146–161, 8 2023.
- [5] R. Batlle-Roca, E. Gómez, W.-H. Liao, X. Serra, and Y. Mitsufuji, “Transparency in music-generative ai: A systematic literature review,” *Research Square* (*preprint*), 12 2024. [Online]. Available: <https://www.researchsquare.com/article/rs-3708077/v2>
- [6] K. Tatar, P. Ericson, K. Cotton, P. T. N. Del Prado, R. Batlle-Roca, B. Cabrero-Daniel, S. Ljungblad, G. Diapoulis, and J. Hussain, “A shift in artistic practices through artificial intelligence,” *Leonardo*, vol. 57, no. 3, pp. 293–297, 2024.
- [7] N. Carlini, J. Hayes, M. Nasr, M. Jagielski, V. Sehwag, F. Tramèr, B. Balle, D. Ippolito, and E. Wallace, “Extracting training data from diffusion models,” in *32nd USENIX Security Symposium (USENIX Security 23)*, 2023, pp. 5253–5270.
- [8] N. Carlini, D. Ippolito, M. Jagielski, K. Lee, F. Tramèr, and C. Zhang, “Quantifying memorization across neural language models,” *arXiv preprint arXiv:2202.07646*, 2023. [Online]. Available: <https://arxiv.org/abs/2202.07646>
- [9] D. Bralios, G. Wichern, F. G. Germain, Z. Pan, S. Khurana, C. Hori, and J. Le Roux, “Generation or replication: Auscultating audio latent diffusion models,” in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 1156–1160.
- [10] O. E. Gundersen, Y. Gil, and D. W. Aha, “On reproducible ai: Towards reproducible research, open science, and digital scholarship in ai publications,” *AI Magazine*, vol. 39, no. 3, pp. 56–68, 2018. [Online]. Available: <https://doi.org/10.1609/aimag.v39i3.2816>
- [11] A. Liesenfeld and M. Dingemanse, “Rethinking open source generative ai: open-washing and the eu ai act,” in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, ser. FAccT ’24. New York, NY, USA: Association for Computing Machinery, 2024, p. 1774–1787. [Online]. Available: <https://doi.org/10.1145/3630106.3659005>
- [12] A. Spirling, “Why open-source generative ai models are an ethical way forward for science,” *Nature*, vol. 616, no. 7957, pp. 413–413, 2023.
- [13] D. Groeneveld, I. Beltagy, P. Walsh, A. Bhagia, R. Kinney, O. Tafjord, A. H. Jha, H. Ivison, I. Magnusson, Y. Wang *et al.*, “Olmo: Accelerating the science of language models,” *arXiv preprint arXiv:2402.00838*, 2024. [Online]. Available: <https://arxiv.org/abs/2402.00838>
- [14] I. Solaiman, “The gradient of generative ai release: Methods and considerations,” in *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, ser. FAccT ’23. New York, NY, USA: Association for Computing Machinery, 2023, p. 111–122. [Online]. Available: <https://doi.org/10.1145/3593013.3593981>

- [15] M. Arnold, R. K. E. Bellamy, M. Hind, S. Houde, S. Mehta, A. Mojsilović, R. Nair, K. N. Ramamurthy, A. Olteanu, D. Piorkowski, D. Reimer, J. Richards, J. Tsay, and K. R. Varshney, “FactSheets: Increasing trust in AI services through supplier’s declarations of conformity,” *IBM Journal of Research and Development*, vol. 63, no. 4/5, pp. 6:1–6:13, 2019.
- [16] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. D. III, and K. Crawford, “Datasheets for datasets,” *Commun. ACM*, vol. 64, no. 12, p. 86–92, nov 2021.
- [17] S. Holland, A. Hosny, S. Newman, J. Joseph, and K. Chmielinski, “The dataset nutrition label,” *Data Protection and Privacy*, vol. 12, no. 12, p. 1, 2020.
- [18] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, and T. Gebru, “Model cards for model reporting,” in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, ser. FAT* ’19. New York, NY, USA: Association for Computing Machinery, 2019, p. 220–229.
- [19] D. R. Desai and M. Riedl, “Between copyright and computer science: The law and ethics of generative ai,” *arXiv preprint arXiv:2403.14653*, 2024. [Online]. Available: <https://arxiv.org/abs/2403.14653>
- [20] Open Source Initiative (OSI), “The Open Source AI Definition – version 1.0,” 10 2024, last accessed: March 26, 2025. [Online]. Available: <https://opensource.org/ai/open-source-ai-definition>
- [21] K. Robison (The Verge), “Open-source AI must reveal its training data, per new OSI definition,” 10 2024, last accessed: March 23, 2025. [Online]. Available: <https://www.theverge.com/2024/10/28/24281820/open-source-initiative-definition-artificial-intelligence-meta-llama>
- [22] R. Gall (The News Stack), “The Open Source AI Definition: What the Critics Say,” 11 2024, last accessed: March 23, 2025. [Online]. Available: <https://thenewstack.io/the-open-source-ai-definition-what-the-critics-say/>
- [23] A. Tarkowski and P. Keller, “The Open Source AI Definition is a step forward in defining openness in AI,” 11 2024, last accessed: March 26, 2025. [Online]. Available: <https://openfuture.eu/blog/the-open-source-ai-definition-is-a-step-forward-in-defining-openness-in-ai/>
- [24] H. H. Jiang, L. Brown, J. Cheng, M. Khan, A. Gupta, D. Workman, A. Hanna, J. Flowers, and T. Gebru, “Ai art and its impact on artists,” in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 2023, pp. 363–374.
- [25] European Parliament and Council of the European Union, “Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act),” 2024, last accessed: March 23, 2025. [Online]. Available: <http://data.europa.eu/eli/reg/2024/1689/oj>
- [26] —, “Article 53: Obligations for Providers of General-Purpose AI Models (Artificial Intelligence Act),” 2024, last accessed: March 23, 2025. [Online]. Available: <https://artificialintelligenceact.eu/article/53/>
- [27] E. A. Office, “General-Purpose AI Models in the AI Act - Questions & Answers,” 03 2025, last accessed: March 26, 2025. [Online]. Available: <https://digital-strategy.ec.europa.eu/en/faqs/general-purpose-ai-models-ai-act-questions-answers>
- [28] EU AI Office, “Second Draft of the General-Purpose AI Code of Practice,” 12 2024, last accessed: March 26, 2025. [Online]. Available: <https://digital-strategy.ec.europa.eu/en/library/second-draft-general-purpose-ai-code-practice-published-written-independent-experts>
- [29] A. Liesenfeld, A. Lopez, and M. Dingemanse, “Opening up chatgpt: Tracking openness, transparency, and accountability in instruction-tuned text generators,” in *Proceedings of the 5th International Conference on Conversational User Interfaces*, ser. CUI ’23. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: <https://doi.org/10.1145/3571884.3604316>
- [30] F. Königstorfer and S. Thalmann, “Ai documentation: A path to accountability,” *Journal of Responsible Technology*, vol. 11, p. 100043, 2022.
- [31] J. Matthews, “Patterns and anti-patterns, principles and pitfalls: accountability and transparency in ai,” *AI Magazine*, vol. 41, no. 1, pp. 82–89, 2020.
- [32] J. Walmsley, “Artificial intelligence and the value of transparency,” *AI & society*, vol. 36, no. 2, pp. 585–595, 2021.
- [33] C. Novelli, M. Taddeo, and L. Floridi, “Accountability in artificial intelligence: what it is and how it works,” *Ai & Society*, vol. 39, no. 4, pp. 1871–1882, 2024.
- [34] R. Bommasani, K. Klyman, S. Longpre, S. Kapoor, N. Maslej, B. Xiong, D. Zhang, and P. Liang, “The foundation model transparency index,” *arXiv preprint arXiv:2310.12941*, 2023. [Online]. Available: <https://doi.org/10.48550/arXiv.2310.12941>
- [35] M. White, I. Haddad, C. Osborne, X.-Y. Liu, A. Abdelmonsef, and S. Varghese, “The model openness framework: Promoting completeness and openness for reproducibility, transparency and usability in ai,” *arXiv preprint arXiv:2403.13784*, 2024. [Online]. Available: <https://doi.org/10.48550/arXiv.2403.13784>

- [36] F. Eiras, A. Petrov, B. Vidgen, C. Schroeder, F. Pizzati, K. Elkins, S. Mukhopadhyay, A. Bibi, A. Purewal, C. Botos, F. Steibel, F. Keshkar, F. Barez, G. Smith, G. Guadagni, J. Chun, J. Cabot, J. Imperial, J. A. Nolasco, L. Landay, M. Jackson, P. H. S. Torr, T. Darrell, Y. Lee, and J. Foerster, “Risks and opportunities of open-source generative ai,” *arXiv preprint arXiv:2405.08597*, 2024. [Online]. Available: <https://doi.org/10.48550/arXiv.2405.08597>
- [37] F. Eiras, A. Petrov, B. Vidgen, C. S. de Witt, F. Pizzati, K. Elkins, S. Mukhopadhyay, A. Bibi, B. Csaba, F. Steibel, F. Barez, G. Smith, G. Guadagni, J. Chun, J. Cabot, J. M. Imperial, J. A. Nolasco-Flores, L. Landay, M. Jackson, P. Röttger, P. H. S. Torr, T. Darrell, Y. S. Lee, and J. Foerster, “Near to mid-term risks and opportunities of open source generative ai,” *arXiv preprint arXiv:2404.17047*, 2024. [Online]. Available: <https://doi.org/10.48550/arXiv.2404.17047>
- [38] J. Engel, K. K. Agrawal, S. Chen, I. Gulrajani, C. Donahue, and A. Roberts, “GANSynth: Adversarial neural audio synthesis,” in *International Conference on Learning Representations (ICLR)*, 2019. [Online]. Available: <https://openreview.net/forum?id=H1xQVn09FX>
- [39] P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford, and I. Sutskever, “Jukebox: A generative model for music,” *arXiv preprint arXiv:2005.00341*, 2020. [Online]. Available: <https://arxiv.org/abs/2005.00341>
- [40] A. Caillon and P. Esling, “Rave: A variational autoencoder for fast and high-quality neural audio synthesis,” *arXiv preprint arXiv:2111.05011*, 2021. [Online]. Available: <https://arxiv.org/abs/2111.05011>
- [41] M. Pasini and J. Schlüter, “Musika! fast infinite waveform music generation,” in *Proceedings of the 23rd International Society for Music Information Retrieval Conference (ISMIR)*, 2022.
- [42] F. Schneider, O. Kamal, Z. Jin, and B. Schölkopf, “Moûsai: Text-to-music generation with long-context latent diffusion,” *arXiv preprint arXiv:2301.11757*, 2023. [Online]. Available: <https://arxiv.org/abs/2301.11757>
- [43] J. Copet, F. Kreuk, I. Gat, T. Remez, D. Kant, G. Synnaeve, Y. Adi, and A. Défossez, “Simple and controllable music generation,” *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, 2024.
- [44] A. Agostinelli, T. I. Denk, Z. Borsos, J. Engel, M. Verzetti, A. Caillon, Q. Huang, A. Jansen, A. Roberts, M. Tagliasacchi *et al.*, “Musiclm: Generating music from text,” *arXiv preprint arXiv:2301.11325*, 2023. [Online]. Available: <https://arxiv.org/abs/2301.11325>
- [45] H. F. Garcia, P. Seetharaman, R. Kumar, and B. Pardo, “Vampnet: Music generation via masked acoustic token modeling,” in *Proceedings of the 24th International Society for Music Information Retrieval Conference (ISMIR)*, 2023.
- [46] K. Chen, Y. Wu, H. Liu, M. Nezhurina, T. Berg-Kirkpatrick, and S. Dubnov, “Musicldm: Enhancing novelty in text-to-music generation using beat-synchronous mixup strategies,” in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 1206–1210.
- [47] S.-L. Wu, C. Donahue, S. Watanabe, and N. J. Bryan, “Music controlnet: Multiple time-varying controls for music generation,” *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 32, p. 2692–2703, May 2024. [Online]. Available: <https://doi.org/10.1109/TASLP.2024.3399026>
- [48] Q. Huang, D. S. Park, T. Wang, T. I. Denk, A. Ly, N. Chen, Z. Zhang, Z. Zhang, J. Yu, C. Frank *et al.*, “Noise2music: Text-conditioned music generation with diffusion models,” *arXiv preprint arXiv:2302.03917*, 2023. [Online]. Available: <https://arxiv.org/abs/2302.03917>
- [49] M. W. Y. Lam, Q. Tian, T. Li, Z. Yin, S. Feng, M. Tu, Y. Ji, R. Xia, M. Ma, X. Song, J. Chen, Y. Wang, and Y. Wang, “Efficient neural music generation,” in *Proceedings of the 37th International Conference on Neural Information Processing Systems*, ser. NIPS ’23. Red Hook, NY, USA: Curran Associates Inc., 2024.
- [50] Z. Novack, J. McAuley, T. Berg-Kirkpatrick, and N. J. Bryan, “DITTO-2: Distilled diffusion inference-time t-optimization for music generation,” in *Proceedings of the 25th International Society for Music Information Retrieval Conference (ISMIR)*, 2024.
- [51] J. Nistal, M. Pasini, C. Aouameur, M. Grachten, and S. Lattner, “Diff-a-riff: Musical accompaniment co-creation via latent diffusion models,” in *Proceedings of the 25th International Society for Music Information Retrieval Conference (ISMIR)*, 2024.
- [52] O. Tal, A. Ziv, I. Gat, F. Kreuk, and Y. Adi, “Joint audio and symbolic conditioning for temporally controlled text-to-music generation,” in *Proceedings of the 25th International Society for Music Information Retrieval Conference (ISMIR)*, 2024.
- [53] Z. Evans, J. D. Parker, C. Carr, Z. Zukowski, J. Taylor, and J. Pons, “Stable audio open,” *arXiv preprint arXiv:2407.14358*, 2024.
- [54] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” in *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*. International Conference on Learning Representations (ICLR), Dec 2014. [Online]. Available: <https://arxiv.org/abs/1312.6114v10>

- [55] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, p. 139–144, Oct. 2020. [Online]. Available: <https://doi.org/10.1145/3422622>
- [56] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
- [57] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems (NeurIPS)*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 6840–6851. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf
- [58] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le, "Flow matching for generative modeling," in *The Eleventh International Conference on Learning Representations*, 2023. [Online]. Available: <https://openreview.net/forum?id=PqvMRDCJT9t>
- [59] C. E. Hill, B. J. Thompson, and E. N. Williams, "A guide to conducting consensual qualitative research," *The Counseling Psychologist*, vol. 25, no. 4, pp. 517–572, 1997. [Online]. Available: <https://doi.org/10.1177/0011000097254001>