MusGU+: Toward a Musician-Centered Evaluation Framework and Discovery Tool for Generative Music AI

Laura Ibáñez-Martínez
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
laura.ibanez@upf.edu

Roser Batlle-Roca
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
roser.batlle@upf.edu

Xavier Serra
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
xavier.serra@upf.edu

Martín Rocamora
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
martin.rocamora@upf.edu

Abstract

Generative music systems are increasingly presented as tools that democratize music creation, yet their practical suitability for musicians remains underexplored. Prior work includes openness-focused evaluation frameworks, such as MusGO (Music-Generative Open AI), as well as qualitative studies of musicians' experiences with generative systems. However, these approaches do not support systematic comparison or early-stage discovery of models for creative use. Motivated by such limitations, we introduce MusGU+, a musician-centered evaluation framework organized around three dimensions: Adaptability, Usability, and Controllability. Together, these capture whether a model can be feasibly trained or fine-tuned on personal data, integrated into real-world music workflows, and controlled in musically meaningful ways. We evaluate 10 representative generative music systems and present an interactive discovery tool that enables musicians to explore and filter models according to these criteria. While MusGO remains valuable for promoting responsible research practices, MusGU+ supports informed selection and practical adoption of generative systems by musicians.

1 INTRODUCTION

Generative music systems have rapidly gained popularity in recent years, particularly through consumer-facing platforms such as Suno (Suno, Inc., 2026) and Udio (Udio, 2026). At the same time, these systems have attracted criticism for their limited transparency regarding training data (Reuters, 2024), as well as broader ethical concerns related to authorship, creative agency, and environmental impact (Barnett, 2023), contributing to growing skepticism toward generative AI. Despite these concerns, such platforms are frequently presented as democratizing music creation (Pram and Morreale, 2025), which raises a fundamental question: *for whom are these systems actually intended?* While platforms like Suno and Udio lower technical barriers and offer more user-friendly interfaces than earlier generative systems, their underlying assumptions about music-making remain at odds with how many musicians understand and value their creative practice. For instance, the CEO of Suno has stated: "the majority of people don't enjoy the majority of the time they spend making music" (Guttridge-Hewitt, 2025), framing generative systems as technologies for optimizing music creation rather than fostering creativity and artistic exploration.

Existing evaluation approaches for generative music systems have mostly relied on output-level evaluation (Agostinelli et al., 2023; Copet et al., 2024; Evans et al., 2025) and, more recently, representation-level analysis (Wei et al., 2024; Ibáñez-Martínez et al., 2026), offering limited insight into how systems can be adapted, controlled, or integrated into real-world musical contexts (Lerch et al., 2025). Complementary approaches have begun to address these limitations from different perspectives. In particular, the Music-Generative Open AI (MusGO) framework (Batlle-Roca et al., 2025) assesses systems in terms of openness and reproducibility. Other work has examined musician-relevant considerations such as data provenance, licensing, creative agency, and user control (Wilson et al., 2025), while practice-oriented studies have further explored how such systems can be incorporated into creative workflows (Dadman et al., 2025; Ronchini et al., 2025). However, these approaches do not provide a systematic comparison of the practical characteristics that shape the suitability of generative systems for musicians, particularly in terms of adaptation, control, and workflow integration.

In this work, we introduce **Music-Generative Usable+ AI (MusGU+)**, a musician-centered evaluation framework designed to address these limitations. MusGU+ adopts the composite and graded evaluation approach followed in MusGO, while shifting the emphasis toward practical suitability for musicians. In particular, it focuses on whether models can be adapted to personal data, integrated into real-world music workflows, and controlled in musically meaningful ways. The framework structures evaluation around three dimensions: **Adaptability**, **Usability**, and **Controllability**, capturing factors such as training pathways, interface availability, workflow integration, and control mechanisms. We apply MusGU+ to 10 generative music systems and present the results through an interactive discovery tool[1] that enables musicians to explore, compare, and filter

[1] https://lauraibnz.github.io/MusGU-plus/

models according to the framework's evaluation criteria[2], supporting more informed selection and practical adoption in creative contexts. As with MusGO, the MusGU+ repository[3] is intended as a living resource, encouraging community engagement through ongoing updates, model additions, and discussion. More broadly, we hope our work encourages further consideration of musician-centered aspects in the development of new generative music systems.

2 BACKGROUND AND RELATED WORK

2.1 Evaluating Generative Music AI

Evaluation of generative music AI has predominantly relied on output-level metrics, particularly for model development and benchmarking. Representative text-to-music systems such as MusicLM (Agostinelli et al., 2023), MusicGen (Copet et al., 2024), and Stable Audio Open (Evans et al., 2025) typically assess generation quality using objective metrics such as Fréchet Audio Distance (FAD) (Kilgour et al., 2019) and Kullback-Leibler (KL) divergence (Kreuk et al., 2022), alongside CLAP-based scores (Wu et al., 2023) for text-prompt adherence. These are often complemented by perceptual listening tests, where participants assess generated samples in terms of realism, coherence, stylistic consistency, and overall preference.

Beyond output-focused evaluation, recent work has begun to analyze the internal representations learned by generative music systems. Such analyses aim to uncover what musical attributes are encoded within these representations, and how this knowledge might support interpretability and controllability. Examples include studies that probe learned representations through music-theoretic lenses (Wei et al., 2024) and approaches that pursue disentanglement to enable controllable music generation (Ibáñez-Martínez et al., 2026). More recent work has also applied mechanistic interpretability methods to identify and manipulate interpretable musical concepts within generative models (Singh et al., 2026). These approaches provide insight into model behavior that is not directly observable from outputs alone.

While output- and representation-level evaluations offer important perspectives on quality and model behavior, they remain limited in accounting for how systems are used and experienced in practice. As noted by Lerch et al. (2025), strong performance on objective metrics or favorable perceptual scores does not necessarily correspond to systems that are usable or controllable in real musical contexts.

2.2 Openness-Focused Evaluation

Among evaluation approaches that consider generative music systems as a whole, a growing focus has been placed on their openness, including source code availability, training data disclosure, licensing, and documentation. The MusGO framework (Batlle-Roca et al., 2025) is a community-driven initiative that assesses generative systems across these dimensions, with the aim of promoting transparency and reproducibility. It also addresses cases of "open-washing", in which models are presented as open despite releasing only a limited subset of components (Liesenfeld and Dingemanse, 2024). MusGO provides a comparative overview through an openness leaderboard, supporting ongoing updates and community contributions.

[2] https://lauraibnz.github.io/MusGU-plus/framework.html

[3] https://github.com/lauraibnz/MusGU-plus

While MusGO acknowledges factors that are relevant to musical practice, such as computational requirements and user-oriented applications, these are not the primary focus of the evaluation. Consequently, models with substantially different levels of practical accessibility or usability may appear similar from an openness perspective. This risks obscuring meaningful differences in the conditions under which systems can be adopted and integrated into real-world musical workflows.

A related perspective grounded in responsible AI principles comes from Wilson et al. (2025), who survey contemporary generative music systems across several musician-relevant dimensions, including input type, real-time generation, licensing, and training or fine-tuning options. While this work raises critical ethical and practical questions, it is primarily a descriptive analysis of the current landscape rather than a structured evaluation. Its focus on musician-relevant concerns highlights considerations that are not systematically reflected in openness-focused evaluations.

2.3 Toward Musician-Centered Perspectives

In parallel, a growing body of literature has examined generative music systems as tools within creative practice, rather than as abstract generators. Several studies have focused on text-to-music models, embedding them in music production settings to study how they support ideation and practical use. Ronchini et al. (2025) report that while text-to-music systems enable rapid exploration, musicians often struggle to translate musical intent into effective control through textual prompts, particularly with respect to temporal structure, arrangement, and expressive detail. Related work (Choi et al., 2025) similarly highlights mismatches between linguistic descriptions and the time-varying, multidimensional nature of musical control.

Further insight comes from workflow-oriented evaluations of generative systems in sound design and music production. Dadman et al. (2025) examine a range of models through hands-on experimentation, identifying challenges related to steerability, latency, configuration complexity, and integration into iterative creative processes. Studies with professional sound designers likewise emphasize the importance of fine-grained and predictable control, noting that opaque or coarse control mechanisms constrain expressive precision and sustained creative use (Kamath et al., 2024). Beyond interaction and control, Bryan-Kinns et al. (2025) examine the role of data scale and system design, highlighting limitations of large-scale training paradigms for non-mainstream musical practices and the potential of small-data approaches to support creative agency.

Other studies adopt a more observational perspective, examining how musicians already incorporate generative AI into their workflows. Pons et al. (2025) document a diverse landscape of AI-assisted music-making across applications such as music production, sound design, and live performance. Their analysis shows that musicians often employ generative models as components within broader creative pipelines, and that working with such systems can be technically demanding, often requiring technical expertise or community support. Thus, despite their growing availability, integration of generative systems into creative workflows remains shaped by both interaction and technical constraints, while existing work offers limited support for systematically comparing models according to musicians' practical needs, particularly in early stages of adoption.

3 MUSGU+: A MUSICIAN-CENTERED EVALUATION FRAMEWORK

3.1 Framework Overview and Design Principles

We introduce the **Music-Generative Usable+ AI (MusGU+)** framework, which adopts a composite and graded evaluation approach inspired by MusGO (Batlle-Roca et al., 2025) and prior openness-focused literature (Liesenfeld and Dingemanse, 2024). Rather than treating evaluation criteria as binary properties, this approach characterizes systems along multiple dimensions, each assessed through clearly defined graded criteria. While MusGO focuses on openness and responsible research practices, MusGU+ shifts toward practical suitability for creative use. Like MusGO, MusGU+ is conceived as a community-oriented living framework, intended to support further refinement, the evaluation of new models, and continued discussion around musician-centered concerns.

The design of MusGU+ draws on prior literature and the authors' experience as researchers in music technology, with a particular focus on generative music systems, including work on controllability and ethical aspects of AI. It also builds on their practice as musicians across different musical contexts, including instrumental performance, music production, and involvement in electronic music scenes.

The framework was iteratively developed through internal discussion and cross-review, and further refined through informal consultation with music producers and sound designers, with the aim of ensuring that its concepts and terminology are understandable and meaningful for practitioners. It was subsequently discussed with participants in the 5th Generative Music AI Workshop[4], held at the Music Technology Group (MTG), providing an additional opportunity for feedback. In this sense, MusGU+ retains a systematic evaluation structure while foregrounding characteristics relevant to musicians and their creative practice.

3.2 Defining Musician-Centered Axes

One motivation for articulating musician-centered evaluation axes emerged from limitations identified in openness-focused criteria, which do not explicitly account for certain practical aspects of model use. For example, RAVE (Caillon and Esling, 2021) supports training on small, user-curated datasets, facilitating adaptation to musicians' own material. While such characteristics are not directly reflected in openness-focused criteria, they have been central to RAVE's use in practice. This aligns with prior work highlighting limitations of large-scale training paradigms and the potential of small-data approaches to support creative agency (Bryan-Kinns et al., 2025). Together, these considerations motivate evaluating the feasibility and available options for adapting models to musicians' own data.

Musician-centered evaluation also requires considering the conditions under which generative systems can be accessed and integrated into musical workflows. RAVE, for instance, supports real-time interaction through environments such as Max/MSP and plugin-based systems (Wilson et al., 2025; Dadman et al., 2025). More broadly, prior work highlights technical and workflow constraints that can shape the practical use of generative systems (Dadman et al., 2025; Pons et al., 2025). This suggests that practical accessibility depends not only on system availability, but also on how readily systems can be run, interacted with, and integrated into creative workflows.

A further motivation concerns controllability, particularly the gap between available control mechanisms and musically meaningful guidance. Prompt-based systems offer limited control over temporal structure and expressive detail (Ronchini et al., 2025; Choi et al., 2025). By contrast, recent approaches such as AFTER (Demerlé et al., 2024), along with earlier work like DDSP (Engel et al., 2020), provide more explicit control over musical attributes, improving interpretability and controllability (Ibáñez-Martínez et al., 2026). In addition, RAVE (Caillon and Esling, 2021) and AFTER (Demerlé et al., 2024) expose internal representations through interactive interfaces, supporting exploratory interaction. These examples highlight the relevance of control mechanisms and interaction possibilities that may support different forms of creative use.

Building on these insights, we organize MusGU+ around three musician-centered evaluation axes:

Adaptability. Assesses how feasible it is for musicians to adapt a model to their own data, including practical constraints and supported training or fine-tuning pathways.

Usability. Examines how easily models can be run, interacted with, and integrated into creative workflows, including access constraints and available support channels.

Controllability. Evaluates the input options and control mechanisms provided by the model, including the diversity, structure, and independence of its control pathways.

3.3 Evaluation Criteria and Structure

The three musician-centered axes are operationalized through 15 evaluation criteria: Adaptability (5), Usability (6), and Controllability (4). A detailed description of the evaluation criteria is provided in Appendix A.

For Adaptability, these criteria include *Hardware Requirements*, *Dataset Size*, *Adaptation Pathways*, *Technical Barriers*, and *Model Redistribution*. Together, these criteria reflect whether a system can be adapted by musicians using realistic computational resources and personal datasets, whether suitable training or fine-tuning pathways are provided, and whether adapted models can be reused or shared beyond the original system.

Usability is evaluated through the criteria *Interface Availability*, *Access Restrictions*, *Real-Time Capabilities*, *Workflow Integration*, *Output Licensing*, and *Community Support*. These criteria assess whether a system can be accessed and run reliably, used interactively, and incorporated into music production or sound design workflows.

Controllability, in turn, includes the criteria *Conditioning Inputs*, *Time-Varying Control*, *Feature Disentanglement*, and *Control Parameters*. These criteria aim to capture the control mechanisms available to musicians and their potential to support intentional musical decisions through precise and predictable guidance of the generation process.

Each criterion is assessed using a three-level graded scale: *fully* (✓), *partially* (∼), and *not supported* (✗), according to the degree to which the corresponding capability is satisfied. In each case, the score levels are accompanied by short descriptive explanations specifying the conditions under which a system would fall into a given category, including representative examples where applicable.

[4] https://www.upf.edu/web/mtg/generative-music-ai-workshop

Table 1: Overview of evaluated generative music systems and rationale for their inclusion in relation to MusGU+ axes.

Model	Year	Architecture	Context	Rationale for Inclusion
AFTER	2024	Latent diffusion, rectified flow	Academic	**Adaptability:** Pretrained audio codec. **Usability:** Real-time; Max/MSP, Max4Live. **Controllability:** Disentangled, time-varying control.
DDSP-VST	2022	Differentiable DSP	Industrial Research	**Adaptability:** Small datasets; custom model training. **Usability:** Real-time; VST/AU. **Controllability:** Time-varying, interpretable control.
JAM	2025	Latent diffusion, rectified flow	Academic	**Adaptability:** "Tiny" architecture. **Controllability:** Word- and phoneme-level timing control.
MusicGen	2023	Autoregressive transformer	Industrial Research	**Adaptability:** Training and fine-tuning pathways. **Controllability:** "Controllable"; time-varying melody conditioning.
Neutone Morpho	2026	Autoencoder	Proprietary	**Adaptability:** Low-barrier custom model training. **Usability:** Real-time; VST/AU.
RAVE	2021	Variational autoencoder	Academic	**Adaptability:** Small datasets; custom model training. **Usability:** Real-time; VST/AU, Max/MSP. **Controllability:** Latent space exploration.
Stable Audio Open Small	2025	Latent diffusion	Industrial Research	**Usability:** "Fast" text-to-audio generation.
Suno	2026	–	Proprietary	**Usability:** Web interface; active Discord community.
Udio	2024	–	Proprietary	**Usability:** Web interface; active Discord community.
YuE	2025	Autoregressive transformer	Academic	**Controllability:** Section-based lyrics-to-song conditioning.

4 MODEL EVALUATION WITH MUSGU+

4.1 Model Selection

We apply the MusGU+ framework to 10 generative music systems in the audio domain: AFTER (Demerlé et al., 2024), DDSP-VST (Google Magenta, 2026), JAM (Liu et al., 2025), MusicGen (Wu et al., 2024), Neutone Morpho (Neutone, 2026), RAVE (Caillon and Esling, 2021), Stable Audio Open Small (Evans et al., 2025), Suno (Suno, Inc., 2026), Udio (Udio, 2026), and YuE (Yuan et al., 2025). An overview of these models is provided in Table 1.

This selection aims to capture a range of current generative music systems that exhibit characteristics relevant to one or more MusGU+ axes and can plausibly be incorporated into real-world music-making workflows. Consequently, previously influential models that are no longer actively maintained, such as Jukebox (Dhariwal et al., 2020) and Musika (Pasini and Schlüter, 2022), are excluded, despite their significance at the time of release. This choice reflects MusGU+'s intention to support early-stage discovery and informed adoption, rather than retrospective evaluation.

The selected models also span diverse architectures and development contexts, covering both academic research and industrial or product-oriented environments, enabling comparison across design goals, intended uses, and musical applications. While not exhaustive, the sample reflects the types of systems MusGU+ is designed to evaluate.

4.2 Evaluation Process

We follow an iterative evaluation process inspired by the methodology used in MusGO. Each model is first evaluated by one of the authors through a review of official resources provided by the developers, including research papers, project websites, documentation, code repositories, and released interfaces. These materials are used to assign scores for each criterion, accompanied by a short justification for every assessment. We also consider supplementary resources acknowledged within the project ecosystem, such as community-developed training notebooks, auxiliary interfaces, or third-party tools referenced by the model developers. Available user interfaces are inspected to assess their functionality and integration pathways, although perceptual or qualitative evaluation of model behaviour and generated audio is beyond the scope of this work.

Each evaluation is then independently reviewed by another author. Discrepancies are discussed and resolved through joint inspection of the evidence, and scores and justifications are refined through one or more iterations until consensus is reached. This cross-review process aims to improve consistency, reduce individual bias, and ensure that evaluations are grounded in verifiable information.

4.3 Interactive Display of Results

The results of the MusGU+ evaluation are shown in Figure 1. Rather than using a static table or leaderboard, we adopt an interactive display that supports exploration and comparison across musical needs and creative priorities. This design choice reflects the view that no single ranking can meaningfully capture the suitability of generative systems across diverse musical contexts, and that evaluation outcomes are most informative when they can be inspected, reordered, and filtered according to intended use.

Aggregate scores are computed for each model by assigning values of 0 for ✗ (*not supported*), 0.5 for ∼ (*partially supported*), and 1 for ✓ (*fully supported*), and summing across criteria and axes. While this aggregation provides a default ordering, it is not intended as a definitive ranking. Instead, the interface allows reordering by individual axes and filtering based on minimum score thresholds (e.g., models with at least 60% support within a given axis).

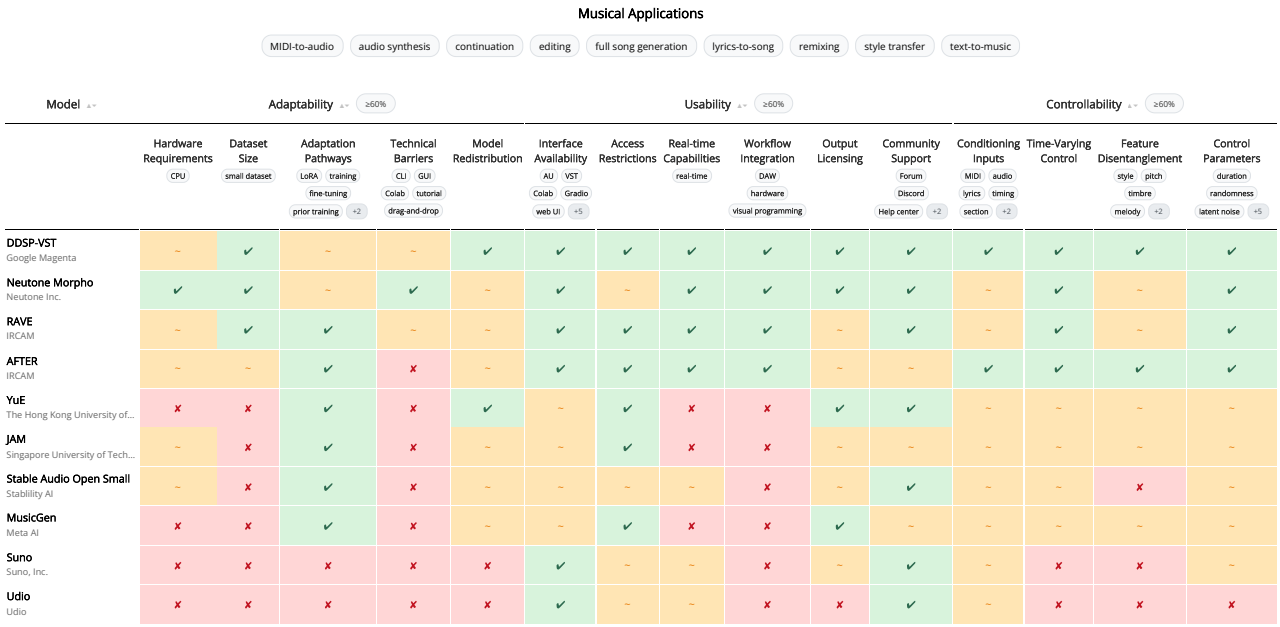

Figure 1: Interactive display of MusGU+ evaluation results for the selected generative music systems across Adaptability, Usability, and Controllability, enabling filtering, reordering, and comparison according to musical needs and priorities.

While aggregate scores provide a useful overview, they do not capture important qualitative distinctions between systems. We therefore introduce reusable tags for selected criteria to support more interpretable and task-oriented comparison. For example, within *Conditioning Inputs*, tags such as *audio*, *MIDI*, and *text prompt* indicate supported input modalities, while within *Workflow Integration*, tags such as *DAW* and *hardware* capture how systems align with established production and performance environments. For other criteria, tags correspond more directly to score levels. In the case of *Real-Time Capabilities*, a single *real-time* tag represents systems that achieve *fully supported* (✓), indicating support for low-latency or interactive use.

The tagging scheme was developed iteratively during the evaluation process, informed by analysis of the selected systems and recurring discussion among the authors. In some cases, this involved abstracting across heterogeneous terminology. For instance, within *Control Parameters*, technical controls (e.g., *temperature*) and more abstract concepts from proprietary systems (e.g., *serendipity*) are mapped to a shared *randomness* tag. In addition to criterion-level tags, we introduce higher-level tags describing *Musical Applications*, such as *style transfer* or *text-to-music*, to support coarse-grained filtering based on typical creative use. The tag set is intentionally provisional and expected to evolve with future model additions and continued community discussion.

5 DISCUSSION

5.1 Insights on the Generative Music AI Landscape

Across the evaluated systems, clear differences emerge. Within Adaptability, only a small number of models, namely DDSP-VST (Google Magenta, 2026), Neutone Morpho (Neutone, 2026), and RAVE (Caillon and Esling, 2021), accommodate the constraints faced by individual musicians, including modest computational requirements, support for small, personally curated datasets, and accessible adaptation pathways. A second group of research-oriented models, including AFTER (Demerlé et al., 2024), YuE (Yuan et al., 2025), JAM (Liu et al., 2025), Stable Audio Open Small (Evans et al., 2025), and MusicGen (Copet et al., 2024), provide formal mechanisms for adaptation but impose significant technical, computational, and data-related barriers, making adaptation difficult in practice. Commercial platforms such as Suno (Suno, Inc., 2026) and Udio (Udio, 2026), in turn, offer no practical pathways for user-driven adaptation. By contrast, Neutone Morpho enables GPU-free model adaptation through a drag-and-drop interface, but imposes access and redistribution restrictions through platform-bound checkpoints.

Regarding Usability, a small group of systems, including DDSP-VST, Neutone Morpho, RAVE, and AFTER, show strong alignment with practical considerations such as interface availability, workflow integration, and real-time interaction, creating a clear gap with the remaining models. By contrast, commercial platforms offer accessible interfaces and strong community support, but often restrict access, output use and ownership. An extreme case is Udio, which prohibits downloading generated outputs. Most of these systems also fail to support integration into music production workflows and rarely enable real-time generation for interactive or performance-oriented contexts.

A similar pattern emerges for Controllability. The same group of systems, including DDSP-VST, Neutone Morpho, RAVE, and AFTER, provide the most extensive support for fine-grained, time-varying control over interpretable musical features, as well as multiple musically meaningful conditioning inputs. These systems also expose a relatively rich set of control parameters, including access to internal representations that can be manipulated for creative exploration. Other models show mixed results, typically offering coarse time-varying control or partial feature disentanglement, with text prompts as the primary conditioning mechanism. Suno and Udio again rank lowest in this axis, providing only global forms of conditioning, largely entangled features and, in Udio's case, no explicit control parameters for shaping generative behavior.

Taken together, the evaluation reveals clear asymmetries across the three axes. Adaptability is the most consistently challenging dimension, often constrained by technical, computational, or platform-bound factors that limit adaptation and redistribution by musicians. Usability shows the greatest variation, with a clear separation between systems designed to integrate into existing music creation workflows and those targeting more general audiences. This challenges democratization claims, suggesting that ease of use does not necessarily translate into meaningful support for musical practice. Controllability occupies a more ambiguous middle ground: many systems offer partial mechanisms for steering generation, yet lack precise, time-localized control over interpretable musical features. Overall, this points to substantial room for improvement in aligning control mechanisms with how musicians conceptualize and shape sound over time.

5.2 A Discovery Tool for Musicians

Beyond presenting evaluation results, the interactive MusGU+ table acts as a discovery tool that enables musicians to explore generative music systems based on concrete creative needs. Broad *Musical Application* tags support filtering for specific purposes, such as text-to-music generation or style transfer. Filtering by evaluation axis (e.g., at least 60% support in Adaptability, Usability, or Controllability) helps identify systems with stronger support along that particular dimension, even when musicians do not yet have clear requirements. Reordering models by individual axes similarly supports comparison of trade-offs across dimensions. At a finer granularity, criterion-level tags enable targeted searches, such as *CPU*-only adaptation under *Hardware Requirements*, *VST* or *Max/MSP* under *Interface Availability*, or explicit *timbre* control under *Feature Disentanglement*.

These interaction mechanisms support exploration across different levels of abstraction, from broad use cases to more specific musical and technical requirements, without assuming a single notion of suitability. In contrast to predominantly research-oriented model listings and benchmarks, MusGU+ foregrounds dimensions that characterize the practical conditions under which generative systems can be adopted in creative practice.

By maintaining MusGU+ as a publicly accessible and evolving resource, we aim to support continued comparison and discussion as new systems emerge. MusGU+ is also intended to engage a broader audience across generative music research and creative practice, encouraging contributions and feedback from both perspectives. More broadly, we hope that articulating and operationalizing these musician-centered dimensions encourages the development of generative music systems that better align with the realities of musical practice.

5.3 Limitations and Future Directions

A key limitation of MusGU+ is the absence of a formal user study or systematic workflow-based evaluation with musicians. Instead, the framework is derived from prior literature, comparative system analysis, and the authors' experience as researchers and musicians. While appropriate for a framework-oriented contribution, this highlights the need for further experiential validation, particularly for Controllability, where future work should examine how available controls behave in practice and support fine-grained musical interaction. By contrast, Adaptability and Usability axes rely more on observable system properties and are therefore less dependent on experiential validation at this stage. MusGU+ also does not aim to capture experiential dimensions such as artistic or aesthetic value, creative support, or potential for collaboration, which require complementary user-centered evaluation. Future work will incorporate workflow-based evaluations with musicians and practitioners, alongside assessment of the framework's usefulness for model discovery and selection.

A second limitation concerns the current scope of the framework. MusGU+ focuses on audio-domain systems for general-purpose music generation and does not yet account for other paradigms, such as symbolic or MIDI-based generation, instrument-specific models, and tools that generate intermediate musical representations. Future work will expand MusGU+ to encompass these modalities and additional models, while incorporating community contributions that help further refine the criteria and assess inter-rater agreement.

Finally, while MusGU+ does not explicitly evaluate ethical or legal dimensions, its musician-centered perspective intersects with broader concerns such as authorship, agency, cultural bias, copyright, and labor impacts, which remain unevenly addressed in the field (Barnett, 2023). Some of these concerns relate to the dimensions considered here: limited controllability can constrain creative agency, while restricted adaptability and access can increase dependency on specific platforms. MusGU+ is therefore not a substitute for ethical assessment, but a complementary lens that foregrounds how system design shapes musicians' engagement with generative systems.

6 CONCLUSION

This paper introduced **MusGU+**, a musician-centered evaluation framework for generative music systems that foregrounds practical suitability for creative use. MusGU+ provides a structured way to evaluate the practical characteristics that shape the adaptability, usability, and controllability of these systems in musical practice.

The framework is organized around three axes: Adaptability, Usability, and Controllability, and operationalized through a graded set of criteria. We apply it to a set of representative generative music systems, revealing persistent asymmetries across these dimensions, particularly in terms of adaptability and musically meaningful control, despite improvements in accessibility. These findings challenge claims of democratization or creative empowerment based primarily on accessibility without considering alignment with musicians' workflows and practical constraints.

We also present an interactive tool that supports exploration and comparison of systems based on musician-centered criteria and tags, enabling informed selection without reducing evaluation to a single ranking. Thus, MusGU+ complements existing research-oriented benchmarks and frameworks by considering dimensions from a musician-centered perspective.

More broadly, MusGU+ contributes a shared vocabulary and evaluative perspective for considering generative music systems as creative tools rather than solely technical artifacts. We hope this work encourages more reflective development practices and supports continued dialogue across technical, creative, and ethical perspectives in the design and adoption of generative music AI systems.

7 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, the European Union-Next Generation EU funds and BMAT Music Innovators, 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, the participants of the 5th Generative Music AI Workshop, and the music producers and sound designers who contributed valuable discussion and feedback that helped refine the framework.

REFERENCES

Agostinelli, A., Denk, T. I., Borsos, Z., Engel, J., Verzetti, M., Caillon, A., Huang, Q., Jansen, A., Roberts, A., Tagliasacchi, M., Sharifi, M., Zeghidour, N., and Frank, C. (2023). MusicLM: Generating music from text. *arXiv preprint arXiv:2301.11325*.

Barnett, J. (2023). The ethical implications of generative audio models: A systematic literature review. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, AIES '23, pages 146–161, New York, NY, USA. Association for Computing Machinery.

Batlle-Roca, R., Ibáñez-Martínez, L., Serra, X., Gómez, E., and Rocamora, M. (2025). MusGO: A community-driven framework for assessing openness in music-generative AI. In *Proceedings of the 26th International Society for Music Information Retrieval Conference (ISMIR)*, pages 727–738, Daejeon, South Korea.

Bryan-Kinns, N., Wszeborowska, A., Sutskova, O., Wilson, E., Perry, P., Fiebrink, R., Viglensoni, G., Lindell, R., Coronel, A., and Correia, N. N. (2025). Leveraging small datasets for ethical and responsible AI music making. In *Proceedings of the 20th International Audio Mostly Conference*, AM '25, pages 70–81, New York, NY, USA. Association for Computing Machinery.

Caillon, A. and Esling, P. (2021). RAVE: A variational autoencoder for fast and high-quality neural audio synthesis. *arXiv preprint arXiv:2111.05011*.

Choi, Y., Moon, J., Yoo, J., and Hong, J.-H. (2025). Understanding the potentials and limitations of prompt-based music generative AI. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA. Association for Computing Machinery.

Copet, J., Kreuk, F., Gat, I., Remez, T., Kant, D., Synnaeve, G., Adi, Y., and Défossez, A. (2024). Simple and controllable music generation. *arXiv preprint arXiv:2306.05284*.

Dadman, S., Bremdal, B. A., and Bergsland, A. (2025). Workflow-based evaluation of music generation systems. *arXiv preprint arXiv:2507.01022*.

Demerlé, N., Esling, P., Doras, G., and Genova, D. (2024). Combining audio control and style transfer using latent diffusion. In *Proceedings of the 25th International Society for Music Information Retrieval Conference (ISMIR)*, pages 721–728, San Francisco, CA, USA.

Dhariwal, P., Jun, H., Payne, C., Kim, J. W., Radford, A., and Sutskever, I. (2020). Jukebox: A generative model for music. *arXiv preprint arXiv:2005.00341*.

Engel, J., Hantrakul, L., Gu, C., and Roberts, A. (2020). DDSP: Differentiable digital signal processing. *arXiv preprint arXiv:2001.04643*.

Evans, Z., Parker, J. D., Carr, C., Zukowski, Z., Taylor, J., and Pons, J. (2025). Stable Audio Open. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, Hyderabad, India.

Google Magenta (2026). DDSP-VST. https://magenta.withgoogle.com/ddsp-vst. Accessed: 2026-02-01.

Guttridge-Hewitt, M. (2025). Suno AI CEO claims people "don't enjoy" making music. https://djmag.com/news/suno-ai-ceo-claims-people-dont-enjoy-making-music. Accessed: 2026-02-01.

Ibáñez-Martínez, L., Nkama, C., Poltronieri, A., Serra, X., and Rocamora, M. (2026). Evaluating disentangled representations for controllable music generation. In *ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain.

Kamath, P., Morreale, F., Bagaskara, P. L., Wei, Y., and Nanayakkara, S. (2024). Sound designer-generative AI interactions: Towards designing creative support tools for professional sound designers. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA. Association for Computing Machinery.

Kilgour, K., Zuluaga, M., Roblek, D., and Sharifi, M. (2019). Fréchet audio distance: A reference-free metric for evaluating music enhancement algorithms. In *Proceedings of Interspeech 2019*, pages 2350–2354. ISCA.

Kreuk, F., Synnaeve, G., Polyak, A., Singer, U., Défossez, A., Copet, J., Parikh, D., Taigman, Y., and Adi, Y. (2022). AudioGen: Textually guided audio generation. *arXiv preprint arXiv:2209.15352*.

Lerch, A., Arthur, C., Bryan-Kinns, N., Ford, C., Sun, Q., and Vinay, A. (2025). Survey on the evaluation of generative models in music. *ACM Computing Surveys*, 58(4).

Liesenfeld, A. and Dingemanse, M. (2024). Rethinking open source generative AI: Open-washing and the EU AI act. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, FAccT '24, pages 1774–1787, New York, NY, USA. Association for Computing Machinery.

Liu, R., Hung, C.-Y., Majumder, N., Gautreaux, T., Bagherzadeh, A. A., Li, C., Herremans, D., and Poria, S. (2025). JAM: A tiny flow-based song generator with fine-grained controllability and aesthetic alignment. *arXiv preprint arXiv:2507.20880*.

Neutone (2026). Neutone Morpho. https://neutone.ai/morpho. Accessed: 2026-02-01.

Pasini, M. and Schlüter, J. (2022). Musika! fast infinite waveform music generation. In Rao, P., Murthy, H., Srinivasamurthy, A., Bittner, R., Repetto, R. C., Goto, M., Serra, X., and Miron, M., editors, *Proceedings of the 23rd International Society for Music Information Retrieval Conference (ISMIR)*, pages 543–550, Bengaluru, India.

Pons, J., Zukowski, Z., Parker, J. D., Carr, C., Taylor, J., and Evans, Z. (2025). Music and artificial intelligence: Artistic trends. *arXiv preprint arXiv:2508.11694*.

Pram, L. and Morreale, F. (2025). Opening musical creativity? embedded ideologies in generative-AI music systems. *arXiv preprint arXiv:2508.08805*.

Reuters (2024). Music labels sue AI companies Suno, Udio for US copyright infringement. https://www.reuters.com/technology/artificial-intelligence/music-labels-sue-ai-companies-suno-udio-us-copyright-infringement-2024-06-24/. Accessed: 2026-02-01.

Ronchini, F., Comanducci, L., Marcucci, S., and Antonacci, F. (2025). AI-assisted music production: A user study on text-to-music models. In *17th International Symposium on Computer Music Multidisciplinary Research (CMMR)*.

Singh, N., Cherep, M., and Maes, P. (2026). Discovering and steering interpretable concepts in large generative music models. In *International Conference on Learning Representations (ICLR)*.

Suno, Inc. (2026). Suno | AI music. https://suno.com. Accessed: 2026-02-01.

Udio (2026). Udio | AI music generator. https://www.udio.com. Accessed: 2026-02-01.

Wei, M., Freeman, M., Donahue, C., and Sun, C. (2024). Do music generation models encode music theory? In *Proceedings of the 25th International Society for Music Information Retrieval Conference (ISMIR)*, pages 680–687, San Francisco, CA, USA.

Wilson, E., Wszeborowska, A., and Bryan-Kinns, N. (2025). A short review of responsible AI music generation. In *Proceedings of the 6th Conference on AI Music Creativity (AIMC)*, Brussels, Belgium.

Wu, S.-L., Donahue, C., Watanabe, S., and Bryan, N. J. (2024). Music ControlNet: Multiple time-varying controls for music generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32:2692–2703.

Wu, Y., Chen, K., Zhang, T., Hui, Y., Berg-Kirkpatrick, T., and Dubnov, S. (2023). Large-scale contrastive language-audio pretraining with feature fusion and keyword-to-caption augmentation. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, Rhode Island, Greece.

Yuan, R., Lin, H., Guo, S., Zhang, G., Pan, J., Zang, Y., Liu, H., Liang, Y., Ma, W., Du, X., Du, X., Ye, Z., Zheng, T., Jiang, Z., Ma, Y., Liu, M., Tian, Z., Zhou, Z., Xue, L., Qu, X., Li, Y., Wu, S., Shen, T., Ma, Z., Zhan, J., Wang, C., Wang, Y., Chi, X., Zhang, X., Yang, Z., Wang, X., Liu, S., Mei, L., Li, P., Wang, J., Yu, J., Pang, G., Li, X., Wang, Z., Zhou, X., Yu, L., Benetos, E., Chen, Y., Lin, C., Chen, X., Xia, G., Zhang, Z., Zhang, C., Chen, W., Zhou, X., Qiu, X., Dannenberg, R., Liu, J., Yang, J., Huang, W., Xue, W., Tan, X., and Guo, Y. (2025). YuE: Scaling open foundation models for long-form music generation. *arXiv preprint arXiv:2503.08638*.

A DETAILED EVALUATION CRITERIA

A.1 Adaptability

Assesses how feasible it is for musicians to adapt a model to their own data, focusing on practical constraints and supported training or fine-tuning pathways.

A.1.1 Hardware Requirements

✓ The model can be trained or fine-tuned on CPU-only systems within a practical timeframe for end users.

∼ Requires a consumer-grade GPU (e.g., T4, P100), such as those found in mid-range gaming laptops or common cloud GPU instances.

✗ Requires dedicated high-end GPUs (e.g., A100s, multi-GPU rigs) or TPUs. Cannot be adapted without access to premium or institutional-level hardware.

A.1.2 Dataset Size

✓ The model can be effectively trained or fine-tuned using small, personal datasets (e.g., minutes to a few hours of audio).

∼ The model requires a significant amount (e.g., tens of hours) of recordings or curated datasets, exceeding the size of most musicians' own libraries.

✗ The model requires large-scale datasets (e.g., hundreds of hours of diverse, high-quality material), making adaptation infeasible without access to institutional or commercial-scale data.

A.1.3 Adaptation Pathways

✓ The model provides practical pathways for training and/or fine-tuning on a musician's own data, such as complete training code, data processing and fine-tuning scripts with required checkpoints, or a dedicated interface for this purpose.

∼ Adaptation pathways are partially available (e.g., incomplete training code, fine-tuning code without checkpoints), requiring users to assemble or infer missing elements.

✗ No practical adaptation pathways are provided. The model cannot be meaningfully trained or fine-tuned using the provided materials.

A.1.4 Technical Barriers

✓ A user-friendly graphical interface or dedicated app is provided for model adaptation, designed for musicians with no programming experience.

∼ The model provides a streamlined setup (e.g., Colab notebook) for training or fine-tuning, with clear instructions and documentation, making it suitable for users with basic technical skills.

✗ Only raw code or scripts are provided, with minimal or no guidance or documentation. Training or fine-tuning requires significant programming knowledge and low-level configuration.

A.1.5 Model Redistribution

✓ Redistribution of trained and/or fine-tuned models or checkpoints is explicitly permitted, allowing adapted models to be shared outside the original system.

∼ Adapted models or checkpoints can be redistributed, but only under constraints (e.g., non-commercial or research-only use, platform-bound sharing).

✗ Redistribution of adapted models or checkpoints is prohibited, contractually restricted, or technically impossible.

A.2 Usability

Examines how easily musicians can run, interact with, and integrate the model into their creative workflows, including access constraints and available support channels.

A.2.1 Interface Availability

✓ A dedicated app or user-friendly graphical interface is provided for running the model (e.g., web platform, HuggingFace Space, standalone GUI, mobile app, Max4Live device), requiring minimal or no setup.

∼ A simplified interface (e.g., Max/MSP or Pure Data patch, Gradio UI code to be run locally) is available but requires some setup or domain-specific familiarity.

✗ Only raw code or scripts are provided for inference, requiring setup from scratch and significant technical knowledge.

A.2.2 Access Restrictions

✓ The model can be used freely and repeatedly without limits, paywalls, or subscriptions. This includes open-source tools with available inference code and pre-trained checkpoints, or publicly accessible platforms with no login requirements or usage limits.

∼ Inference is possible, but access is constrained (e.g., login required, acceptance of terms of use, daily usage limits, or restricted free tiers), introducing account-based or usage limitations.

✗ Direct inference is not feasible due to missing components (e.g., no pretrained checkpoint or inference code), or access is fully restricted (e.g., paywalls, subscriptions, or unstable user interfaces).

A.2.3 Real-Time Capabilities

✓ The system generates output with minimal or imperceptible delay (e.g., milliseconds to a few seconds), making it suitable for live use even on modest hardware.

∼ Generation is possible with a moderate delay (e.g., several seconds), making it usable in interactive settings but not for live use. Real-time performance may require a consumer-grade GPU.

✗ Generation is slow (e.g., minutes per sample) or impractical for real-time use, especially on typical personal computers or laptops.

A.2.4 Workflow Integration

✓ The model is directly usable in common music workflows (e.g., within DAWs, visual programming environments, live coding systems, or dedicated music hardware).

∼ Some integration is possible (e.g., via OSC/MIDI connectivity or similar control-based interfaces).

✗ The system is isolated, with no clear path to embed it in existing creative setups or musical instruments.

A.2.5 Output Licensing

✓ The output is fully usable for both personal and commercial purposes without restriction.

∼ Some use limitations apply to the generated output (e.g., non-commercial use only, attribution required, unclear terms).

✗ Output use is heavily restricted or prohibited (e.g., proprietary licensing, unclear or forbidding terms).

A.2.6 Community Support

✓ An open, user-facing community space is available (e.g., Discord server or forum), suitable for musicians to ask questions and share workflows.

∼ Limited or developer-oriented channels are available (e.g., GitHub issues), which may provide assistance but are less accessible to most musicians.

✗ No meaningful public support or community spaces are provided.

A.3 Controllability

Evaluates the kinds of input and internal control mechanisms a model offers to guide its behavior, including the diversity, structure, and independence of its control pathways.

A.3.1 Conditioning Inputs

✓ The model accepts multiple and diverse conditioning inputs, including at least two musically meaningful modalities (e.g., audio, MIDI, symbolic), enabling rich and varied guidance during generation.

∼ Conditioning is limited to one musically meaningful modality, possibly combined with descriptive inputs (e.g., melody and text).

✗ Offers little to no input conditioning. Generation is mostly uncontrolled or limited to coarse or global labels (e.g., vibe, energy).

A.3.2 Time-Varying Control

✓ Supports precise time-varying control (e.g., per-frame pitch, dynamics, or symbolic structure), enabling fine-grained and expressive manipulation.

∼ Some time-localized control is available (e.g., melody following or segment-based prompts), but not fine-grained.

✗ Only global control is possible, with no ability to influence generation over time.

A.3.3 Feature Disentanglement

✓ Provided controls are designed to influence distinct and isolated musical attributes (e.g., timbre, pitch, rhythm, structure), enabling predictable and interpretable guidance.

∼ Some effort is made to separate control pathways, but conditioning inputs are not explicitly associated with interpretable musical attributes, and interactions are expected.

✗ Control signals are entangled or loosely defined, making it unclear how specific inputs relate to individual musical attributes.

A.3.4 Control Parameters

✓ Provides multiple configurable model parameters (e.g., duration, randomness, conditioning strength) and enables direct manipulation of internal representations.

∼ Provides a limited set of configurable parameters and/or restricted access to latent representations.

✗ No additional meaningful parameters or internal controls are exposed beyond the primary conditioning mechanisms, if any.