Live Improvisation with Fine-Tuned Generative AI: A Musical Metacreation Approach

Misagh Azimi
Victoria University of Wellington
PO Box 600
Wellington 6140,
New Zealand
misagh.azimi@vuw.ac.nz

Mo H. Zareei
Victoria University of Wellington
PO Box 600
Wellington 6140,
New Zealand
mo.zareei@vuw.ac.nz

ABSTRACT
This paper presents a pipeline to integrate a fine-tuned open-source text-to-audio latent diffusion model into a workflow with Ableton Live for the improvisation of contemporary electronic music. The system generates audio fragments based on text prompts provided in real time by the performer, enabling dynamic interaction. Guided by Musical Metacreation as a framework, this case study reframes generative AI as a co-creative agent rather than a mere style imitator. By fine-tuning Stable Audio Open on a dataset of the first author's compositions and field recordings, this approach demonstrates the ethical and practical benefits of open-source solutions. Beyond showcasing the model's creative potential, this study highlights the model's significant challenges and the need for democratized tools with real-world applications.

Author Keywords
Generative AI, Improvisation, Real-time Interaction, Musical Metacreation, Democratizing AI

1. INTRODUCTION
Recent years have seen a surge in text-to-audio and text-to-music research [1, 12–14, 16, 41, 45], driven by advances in machine learning that promise novel sonic possibilities [6, 8]. Generative AI (GenAI) models now cover a wide range of genres, while non-generative AI is used to automate tasks such as mastering [28, 34]. Many text-to-audio models that generate polished compositions are reshaping traditional musical practices [37, 38], potentially devaluing musicians' work [29]. Moreover, many of these models are trained on copyrighted music [36].

Yet, the practical impact of AI on music-making remains underexamined [4, 33]. Earlier AI music research often focused on style replication [2], although recent work increasingly emphasizes originality and collaborative interaction [10, 15, 21, 27]. To explore these evolving priorities, we propose a pipeline that integrates a fine-tuned text-to-audio diffusion model into live performance, thereby exploring GenAI's potential in real-time improvisation. Building on traditions such as live coding [5, 7, 9, 26] and algorithmic composition [30], our approach extends prior AI-related NIME research [3, 19, 24, 40, 43]. We emphasize diffusion-based generation over other methods (e.g., rule-based, symbolic) and deploy the resulting pipeline in a live improvisational setting.

1.1 From Demonstration to Democratic Intervention
Sturm et al. [33] , building on Wagstaff's position [42], argue that demonstration-driven metrics like perplexity or human-like simulation often lack relevance for musicians. As Ben-Tal et al. [4] note, evaluating AI music models requires assessing their real-world usefulness in composition, collaboration, and improvisation.

Furthermore, Feenberg [18] argues that user appropriation can serve as a democratic intervention by enabling practitioners to shape emerging technologies rather than merely adopt them. This ethos resonates with practices like circuit bending and hacking [11, 20, 25], all of which involve retooling existing technologies to meet creative needs.

Implementing this vision towards GenAI centers on open-source, ethically trained models. Unlike proprietary models [1, 14, 35, 39] or "open" ones using privately licensed data [12], Stable Audio Open [17] uses only Creative Commons licensed audio, aligning with democratizing ideals—despite possible quality trade-offs. We, therefore, adopt it in our case study, prioritizing access and creative freedom.

1.2 Key Challenges
Integrating text-to-audio models into real-world musical environments presents several significant challenges:

- *Proprietary Models:* Many leading models remain proprietary, limiting fine-tuning and experimentation [17, 32].
- *Data Origin:* Web-scraped audio raises licensing concerns that can discourage open adoption [38].
- *Open-Model Shortcomings:* Open-source models often lag in quality and coherence compared to commercial models [17].
- *Computational Demand:* Using and fine-tuning generative models requires substantial computational resources [23, 44, 46], often out of reach for many.

With these challenges outlined, we now focus on the frameworks informing the design of our system.

2. MUSICAL METACREATION AS FRAMEWORK
Musical Metacreation (MuMe), a subset of computational creativity, explores the interplay between generative autonomy and human adaptation, providing the conceptual foundation for this study. Pasquier et al. [31] describe MuMe systems as capable of generating, transforming, or analyzing musical content with varying degrees of autonomy.

In improvised performances, MuMe systems can supply musical fragments—melodic snippets or textures—that performers deploy. This collaborative interaction emphasizes the system's generative creativity (producing novel sound artifacts) and the performer's adaptive creativity (integrating and reshaping artifacts within the performance) [31].

Their framework informs our approach through three key aspects: an autonomy continuum that balances system-driven and human-guided processes; a distinction between corpus-based and non-corpus-based approaches—where the latter generates outputs without being exposed to musical information as input, determining contextual novelty; and a classification of systems as operating in real time versus offline to guide analyses of interactive collaboration.

NIME'25, June 24–27, 2025, Canberra, Australia

Within this context, we position GenAI as a collaborative agent[1]. In our live improvisation pipeline, GenAI generates musical fragments from real-time prompts, shifting the focus from problem-solving (optimal outcomes) to "problem-seeking" [2, 22], enabling musicians to reshape or subvert the outputs creatively.

3. SYSTEM DESIGN AND IMPLEMENTATION

The proposed pipeline operates in three consecutive stages: (1) Conditioning the AI model (prompting), (2) Generating fragments (outputs), and (3) Integrating these fragments in Ableton. Iterating these steps aims to create a feedback loop between the performer and AI, balancing model autonomy and human creativity. The following subsections detail the text-to-audio model, fine-tuning process, and Ableton integration.

3.1 The Fine-Tuned Latent Diffusion Model

Stable Audio Open 1.0 is an open-source text-to-audio model trained on 7,300 hours of Creative Commons licensed audio and music, capable of generating 44.1 kHz stereo outputs up to 47 seconds. To further align the system with the concept of user appropriation, this model was fine-tuned using music and audio sourced from the first author's repertoire, aiming to reflect their aesthetic preferences in the generated outputs.

3.1.1 Dataset Preparation for Fine-Tuning

A dataset of approximately 41 hours of music, improvisations, sketches, and field recordings was segmented into 5-second chunks following Evans et al.'s original method [17]. After removing silent and short chunks, this yielded 23,003 files (.wav), corresponding to roughly 32 hours of audio and music.

A structured metadata tagging strategy was adopted for fine-tuning: full compositions were manually annotated with attributes (title, genre, mood, tempo, key, description, name), and their individual instrument stems inherited these attributes, adding unique details (e.g., instrument names) extracted from filenames.

3.1.2 Training the Model

Training used the described audio dataset with metadata to strengthen links between text prompts and audio outputs. The pre-trained checkpoint available on HuggingFace[2] was used with configuration settings from the stable-audio-tools[3] library. Training was conducted on an NVIDIA A100 in Google Colab, and the resulting checkpoint functions as the fine-tuned model for our improvisational pipeline.

3.2 Text-to-Audio Generation

To generate audio, the model is conditioned with short natural language prompts entered through a simple interface featuring a text field and a generation button (Figure 1).

Figure 1. Gradio UI to enter prompts

The system loads the fine-tuned checkpoint at startup using the stable-audio-tools library configuration and generates audio fragments when the Generate button is pressed. Inference time can vary depending on the GPU e.g. 7 seconds on an RTX 4090 versus 12 minutes on a MacBook Pro M1 (Figure 2).

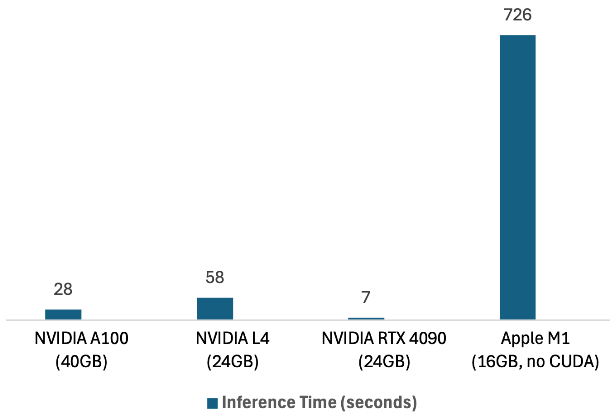

Figure 2. Inference times using different GPUs

Generated fragments are saved as .wav files in a four-file loop: "Fragment01.wav" to "Fragment04.wav. " New outputs overwrite the oldest file in a sequence of four, allowing performers to improvise with four simultaneous tracks after they are imported into Ableton, as shown in the next step.

3.3 Integration into Ableton Live

A custom Max for Live patch imports AI-generated fragments into Ableton, where each can be buffered and triggered on demand. These fragments serve as immediate raw material for improvisation—ready to be played, looped, or manipulated like any other element. Inserted into an Ableton track, the interface (Figure 3) lets users select from four fragments (1–4).

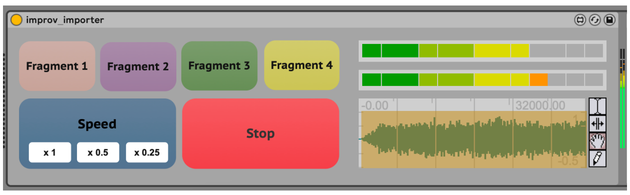

Figure 3. Max for Live patch to buffer and play fragments

Since the generated fragments are relatively short (≤47 seconds), half- and quarter-time-stretch functions were integrated to extend their duration if desired. Additionally, all interactive elements of the patch are MIDI-mapped through Ableton, enabling tangible, haptic control.

3.4 Live Improvisation

The concluding pipeline (Figure 4) includes the following steps:

- 1. *Initiation:* The system loads the model checkpoint,
- 2. *Conditioning:* The performer prompts the model,
- 3. *Generation:* The model generates an audio fragment,
- 4. *Import:* The fragment is buffered and imported into Ableton,
- 5. *Improvisation:* The performer engages with the fragment,
- 6. *Iteration:* Repeating the process to extend the improvisation.

[1] A MuMe agent is a system designed to autonomously perform creative musical tasks (e.g. composition or improvisation) and can operate both online and offline. Its level of autonomy ranges from fully generative systems to interactive, computer-assisted tools.

[2] https://huggingface.co/stabilityai/stable-audio-open-1.0
[3] https://github.com/Stability-AI/stable-audio-tools

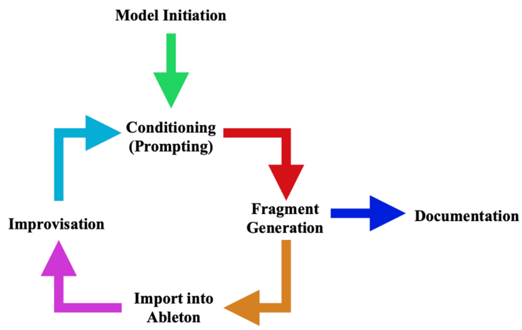

Figure 4. The proposed improvisation pipeline

In practice[4], the performer only needs to input a prompt and trigger generation through the UI, while all musical actions—playback and manipulation—are handled in real time via a MIDI controller (Figure 5).

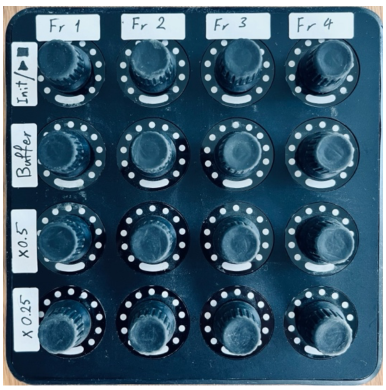

Figure 5. MIDI-mapped parameters for haptic interaction

4. OBSERVATIONS FROM SYSTEM USE

Live deployment revealed technical challenges but also substantial creative potential. This section summarizes key observations from real-world pipeline use, highlighting how a fine-tuned text-to-audio model can both support and challenge the performer's creative process.

4.1 Technical Challenges

During the case study we faced numerous challenges. Training was resource-intensive (>33 GB VRAM; Figure 6) and unreliable in Colab due to automatic updates and lack of an isolated environment, leading to dependency issues (e.g., NumPy and PyTorch), immensely affecting reproducibility. We partially addressed this by manually installing libraries and reverting to compatible versions, though this was not always effective. We recommend running the process locally when possible.

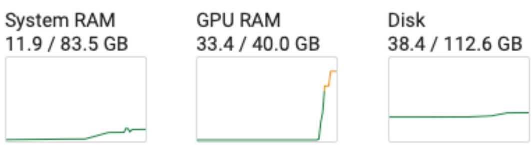

Figure 6. Computational demand during training on Colab

In addition, some generated fragments exhibited issues such as long silent parts (Figure 7) and tempo variations, which could disrupt the musical flow when looped. Handling tempo variations—e.g. quantizing fragments to a strict tempo grid with low latency and no human intervention—was challenging and fell outside the study's scope. Occasional silences, however, proved manageable with a lightweight post-processing script[5].

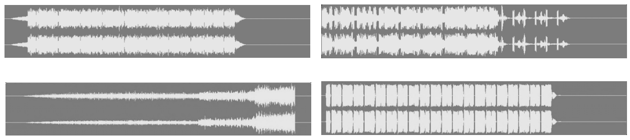

Figure 7. Some generated fragments exhibit silent sections

The fine-tuning process relied on several hours of audio from the first author's repertoire, introducing two further reproducibility challenges. Firstly, other artists or researchers will not have access to such data, and the impact of fine-tuning can be limited even with ample data. Secondly, we cannot publish the fine-tuning dataset as open-source due to limitations of copyright, co-ownership, and contractual agreements. Since this was a case study and not the training of a foundation model, using this dataset was considered permissible within the context of publishing the findings. Open-sourcing this data, however, raises legal concerns beyond the scope of this case study.

4.2 Creative Potential

Despite limitations, the system exhibits notable creative potential. While tempo variations challenge improvisation in genres that rely on steady beats (e.g., pop, EDM), the model excels at generating fragments suited to more experimental styles. As Evans et al. [16] observed, its strength lies in sound design rather than in producing polished compositions—a quality that opens exciting opportunities for subversive use in experimental music. The model's unexpected sonic outputs prompt performers to adapt and explore new musical directions, acting as a creative partner that provokes innovative ideas without undermining human agency. Additionally, the system's high degree of autonomy combined with aesthetic characteristics reflective of the author's musical repertoire offers a fresh approach to improvisation.

5. SYSTEM EVALUATION

The subjectivity of musical creativity makes evaluating MuMe systems challenging [31]. With that in mind, we evaluate our system using criteria proposed by Pasquier et al. [31] alongside additional MuMe aspects:

- *Creativity:* Once conditioned, the system autonomously generates musical material while performers refine and adapt the output in real-time—demonstrating exploratory, generative, and adaptive forms of creativity.
- *Contextual Novelty:* Trained on our work yet receiving no audio input during improvisation, this non-corpus-based system generates novel content within a defined aesthetic.
- *Quality and Reliability:* The output generally showcased useful, unexpected, and engaging textures, although some were subjectively rated as unsatisfactory by the authors.
- *Robustness:* The system produces output on all input prompts.
- *Interactive Collaboration*: A feedback loop between prompt, AI output, and live manipulation supports MuMe's vision of systems that inspire and respond to human creativity.

[4] A video of real-time interaction with the system can be viewed here: https://misaghazimi.com/research/sao-improv

[5] https://github.com/MAz-Codes/sao-silence-remover.git

By evaluating the system along these dimensions, we situate its contributions within the broader discourse of MuMe.

6. CONCLUSION AND FUTURE WORK

Our case study shows GenAI can extend musical creativity beyond mere imitation. Rapid, responsive sound generation enables AI to catalyze new ideas and open fresh improvisational pathways. Yet despite this potential, high computational and data demands may hinder broader adoption.

Future work will include evaluations with audiences and artists to assess artistic impact and guide improvements. Additionally, fine-tuning more effectively with smaller datasets can reduce data needs and expand accessibility. Developing optimized pipelines that run on local hardware with limited resources will also be crucial.

As this technology evolves and opens new possibilities for Musical Metacreation, our research aims to lay the groundwork for deeper human-AI collaboration in live improvisation.

7. ACKNOWLEDGMENTS

Our thanks to Dr. Tim Corballis for his invaluable input and insightful feedback.

8. ETHICAL STANDARDS

The model used in this case study was trained exclusively on Creative Commons–licensed audio (CC-0, CC-BY, or CC-Sampling+). The dataset used for fine-tuning consists solely of audio owned or co-owned by the first author. The authors declare that no conflicts of interest are associated with this research.

9. REFERENCES

[1] Andrea Agostinelli, Timo I. Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank. 2023. MusicLM: Generating Music From Text. https://doi.org/10.48550/arXiv.2301.11325

[2] Christopher Ariza. 2009. The Interrogator as Critic: The Turing Test and the Evaluation of Generative Music Systems. *Computer Music Journal* 33, 2 (2009), 48–70.

[3] Christodoulos Benetatos, Joseph VanderStel, and Zhiyao Duan. 2020. BachDuet: A Deep Learning System for Human-Machine Counterpoint Improvisation. In. (Proceedings of the International Conference on New Interfaces for Musical Expression). 2020. 635–640. https://doi.org/10.5281/zenodo.4813234

[4] Oded Ben-Tal, Matthew Tobias Harris, and Bob L. Sturm. 2021. How Music AI Is Useful: Engagements with Composers, Performers and Audiences. *Leonardo* 54, 5 (October 2021), 510–516. https://doi.org/10.1162/leon_a_01959

[5] Alan F. Blackwell, Emma Cocker, Geoff Cox, Alex McLean, and Thor Magnusson. 2022. *Live Coding: A User's Manual*. The MIT Press. https://doi.org/10.7551/mitpress/13770.001.0001

[6] Jean-Pierre Briot, Gaëtan Hadjeres, and François-David Pachet. 2020. *Deep Learning Techniques for Music Generation*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-70163-9

[7] Andrew R. Brown and Andrew Sorensen. 2009. Interacting with Generative Music through Live Coding. *Contemporary Music Review* 28, 1 (February 2009), 17–29. https://doi.org/10.1080/07494460802663991

[8] Filippo Carnovalini and Antonio Rodà. 2020. Computational Creativity and Music Generation Systems: An Introduction to the State of the Art. *Frontiers in Artificial Intelligence* 3, (2020). Retrieved March 4, 2024 from https://www.frontiersin.org/articles/10.3389/frai.2020.00014

[9] Nick Collins, Alex McLEAN, Julian Rohrhuber, and Adrian Ward. 2003. Live coding in laptop performance. *Organised Sound* 8, 3 (December 2003), 321–330. https://doi.org/10.1017/S135577180300030X

[10] Nick Collins, Vit Ruzicka, and Mick Grierson. 2020. Remixing AIs: mind swaps, hybrainity, and splicing musical models. (December 2020). Retrieved April 15, 2025 from https://durham-repository.worktribe.com/output/1139583/remixing-ais-mind-swaps-hybrainity-and-splicing-musical-models

[11] Nicolas Collins. 2020. *Handmade electronic music: the art of hardware hacking* (Third edition. ed.). Routledge, New York.

[12] Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez. 2024. Simple and Controllable Music Generation. https://doi.org/10.48550/arXiv.2306.05284

[13] DeepMind. 2023. Transforming the future of music creation. *Google DeepMind*. Retrieved April 22, 2024 from https://deepmind.google/discover/blog/transforming-the-future-of-music-creation/

[14] Prafulla Dhariwal, Heewoo Jun, Christine Payne, Jong Wook Kim, Alec Radford, and Ilya Sutskever. 2020. Jukebox: A Generative Model for Music. https://doi.org/10.48550/arXiv.2005.00341

[15] Arne Eigenfeldt. 2022. Musebots and I: Collaborating with Creative Systems. In *The Language of Creative AI: Practices, Aesthetics and Structures*, Craig Vear and Fabrizio Poltronieri (eds.). Springer International Publishing, Cham, 207–216. https://doi.org/10.1007/978-3-031-10960-7_12

[16] Zach Evans, C. J. Carr, Josiah Taylor, Scott H. Hawley, and Jordi Pons. 2024. Fast Timing-Conditioned Latent Audio Diffusion. https://doi.org/10.48550/arXiv.2402.04825

[17] Zach Evans, Julian D. Parker, C. J. Carr, Zack Zukowski, Josiah Taylor, and Jordi Pons. 2024. Stable Audio Open. https://doi.org/10.48550/arXiv.2407.14358

[18] Andrew Feenberg. 1991. *Critical theory of technology*. Oxford University Press, New York.

[19] Rebecca Fiebrink, Dan Trueman, and Perry R. Cook. 2009. A Meta-Instrument for Interactive, On-the-Fly Machine Learning. In. (Proceedings of the International Conference on New Interfaces for Musical Expression). 2009. 280–285. https://doi.org/10.5281/zenodo.1177513

[20] Qubais Reed Ghazala. 2021. Circuit-Bending and Living Instruments. In *Sound Inventions: Selected Articles from Experimental Musical Instruments* (1st ed.), Bart Hopkin and Sudhu Tewari (eds.). Taylor & Francis Group, Oxford, UNITED KINGDOM. Retrieved September 10, 2024 from http://ebookcentral.proquest.com/lib/vuw/detail.action?docID=6680914

[21] Guy Hoffman and Gil Weinberg. 2010. Shimon: an interactive improvisational robotic marimba player. In. *CHI '10 Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA, April 10,2010. Association for Computing Machinery, New York, NY, USA, 3097–3102. https://doi.org/10.1145/1753846.1753925

[22] Simon Holland. 2020. Artificial Intelligence in Music Education: A Critical Review. In *Readings in Music and Artificial Intelligence*, Eduardo Reck Miranda (ed.). Routledge.

[23] Andre Holzapfel, Anna-Kaisa Kaila, and Petra Jääskeläinen. 2024. Green MIR? Investigating Computational Cost of Recent Music-Ai Research in ISMIR. ISMIR, San Francisco, California, USA and Online, 371–380. https://doi.org/10.5281/zenodo.14877351

[24] Atsuya Kobayashi, Ryo Nishikado, and Nao Tokui. 2023. Improvise+=Chain: Listening to the Ensemble Improvisation of an Autoregressive Generative Model. In. (Proceedings of the International Conference on New Interfaces for Musical Expression). 2023. 633–636. https://doi.org/10.5281/zenodo.11189329

[25] Steven Levy. 2010. *Hackers: heroes of the computer revolution* (First edition. ed.). O'Reilly Media, Beijing ;

[26] Thor Magnusson. 2011. Algorithms as Scores: Coding Live Music. *Leonardo Music Journal* 21, (December 2011), 19–23. https://doi.org/10.1162/LMJ_a_00056

[27] Gianluca Micchi, Louis Bigo, Mathieu Giraud, Richard Groult, and Florence Levé. 2021. I Keep Counting: An Experiment in Human/AI Co-creative Songwriting. 4, 1 (December 2021), 263–275. https://doi.org/10.5334/tismir.93

[28] Eduardo Reck Miranda (Ed.). 2021. *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-72116-9

[29] Fabio Morreale. 2021. Where Does the Buck Stop? Ethical and Political Issues with AI in Music Creation. 4, 1 (July 2021), 105–113. https://doi.org/10.5334/tismir.86

[30] Gerhard Nierhaus. 2009. *Algorithmic Composition: Paradigms of Automated Music Generation* (1st ed.). Springer Vienna. Retrieved April 11, 2024 from https://link.springer.com/book/10.1007/978-3-211-75540-2

[31] Philippe Pasquier, Arne Eigenfeldt, Oliver Bown, and Shlomo Dubnov. 2017. An Introduction to Musical Metacreation. *Comput. Entertain.* 14, 2 (January 2017), 2:1-2:14. https://doi.org/10.1145/2930672

[32] Irene Solaiman. 2023. The Gradient of Generative AI Release: Methods and Considerations. In. *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. New York, NY, USA, June 12,2023. Association for Computing Machinery, New York, NY, USA, 111–122. https://doi.org/10.1145/3593013.3593981

[33] Bob L. Sturm, Oded Ben-Tal, Úna Monaghan, Nick Collins, Dorien Herremans, Elaine Chew, Gaëtan Hadjeres, Emmanuel Deruty, and François Pachet. 2019. Machine learning research that matters for music creation: A case study. *Journal of New Music Research* 48, 1 (January 2019), 36–55. https://doi.org/10.1080/09298215.2018.1515233

[34] Bob L. Sturm, Maria Iglesias, Oded Ben-Tal, Marius Miron, and Emilia Gómez. 2019. Artificial Intelligence and Music: Open Questions of Copyright Law and Engineering Praxis. *Arts* 8, 3 (September 2019), 115. https://doi.org/10.3390/arts8030115

[35] Suno. 2024. Suno. Retrieved April 22, 2024 from https://suno.com/

[36] Suno, Inc. 2024. *Answer of Defendant Suno, Inc. to Complaint*. United States District Court for the District of Massachusetts.

[37] Koray Tahiroğlu. 2021. Ever-shifting roles in building, composing and performing with digital musical instruments. *Journal of New Music Research* 50, 2 (March 2021), 155–164. https://doi.org/10.1080/09298215.2021.1900275

[38] Kivanç Tatar, Petter Ericson, Kelsey Cotton, Paola Torres Núñez Del Prado, Roser Batlle-Roca, Beatriz Cabrero-Daniel, Sara Ljungblad, Georgios Diapoulis, and Jabbar Hussain. 2024. A Shift in Artistic Practices through Artificial Intelligence. *Leonardo* 57, 3 (June 2024), 293–297. https://doi.org/10.1162/leon_a_02523

[39] Udio. 2024. Udio | Make your music. *Udio*. Retrieved April 22, 2024 from https://udio.com

[40] Satvik Venkatesh, Edward Braund, and Eduardo Miranda. 2020. Composing Popular Music with Physarum polycephalum-based Memristors. In. (Proceedings of the International Conference on New Interfaces for Musical Expression). 2020. 514–519. https://doi.org/10.5281/zenodo.4813507

[41] Apoorv Vyas, Bowen Shi, Matthew Le, Andros Tjandra, Yi-Chiao Wu, Baishan Guo, Jiemin Zhang, Xinyue Zhang, Robert Adkins, William Ngan, Jeff Wang, Ivan Cruz, Bapi Akula, Akinniyi Akinyemi, Brian Ellis, Rashel Moritz, Yael Yungster, Alice Rakotoarison, Liang Tan, Chris Summers, Carleigh Wood, Joshua Lane, Mary Williamson, and Wei-Ning Hsu. 2023. Audiobox: Unified Audio Generation with Natural Language Prompts. https://doi.org/10.48550/arXiv.2312.15821

[42] Kiri Wagstaff. 2012. Machine Learning that Matters. https://doi.org/10.48550/arXiv.1206.4656

[43] Ian Whalley. 2010. Generative Improv . & Interactive Music Project (GIIMP). In. (Proceedings of the International Conference on New Interfaces for Musical Expression). 2010. 255–258. https://doi.org/10.5281/zenodo.1177923

[44] Yuchen Xia, Jiho Kim, Yuhan Chen, Haojie Ye, Souvik Kundu, Cong Hao, and Nishil Talati. 2024. Understanding the Performance and Estimating the Cost of LLM Fine-Tuning. https://doi.org/10.48550/arXiv.2408.04693

[45] Yao Yao, Peike Li, Boyu Chen, and Alex Wang. 2024. JEN-1 Composer: A Unified Framework for High-Fidelity Multi-Track Music Generation. https://doi.org/10.48550/arXiv.2310.19180

[46] Zongyu Yin, Federico Reuben, Susan Stepney, and Tom Collins. 2023. Deep learning's shallow gains: a comparative evaluation of algorithms for automatic music generation. *Mach Learn* 112, 5 (May 2023), 1785–1822. https://doi.org/10.1007/s10994-023-06309-w

10. Appendix

A video demonstrating the system's real-time utilization alongside some audio examples of improvisation sessions can be found here: https://misaghazimi.com/research/sao-improv