Longevity of Deep Generative Models in NIME: Challenges and Practices for Reactivation

Isaac Joseph Clarke
CMA, Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
ijclarke590@connect.hkust-gz.edu.cn

Francesco Ardan Dal Rí
Dept. of Information Engineering and Computer Science
University of Trento, Italy
francesco.dalri-2@unitn.it

Raul Masu
Conservatorio F.A. Bonporti
Trento, Italy; and
CMA, Hong Kong University of Science and Technology (Guangzhou), China
raul@raulmasu.org

Abstract

In this paper, we present an investigation into the longevity, reproducibility, and documentation quality of Deep Generative Models (DGMs) introduced in previous editions of the NIME conference. We begin by assessing whether DGM presented at NIME are still available in terms of code, data, and weights; afterward, we present the recreation process of seven unavailable models, to the end of investigation of the issues related to longevity and documentation. We examine the availability and completeness of resources needed to recreate DGM models, and discuss specific challenges encountered during such recreation. Drawing from this experience, we highlight key obstacles that hinder the long-term viability and reuse of DGMs in the NIME context, and propose guidelines to improve their documentation and future reuse within the community.

Keywords

Longevity, Reuse, Recreation, Documentation, Sustainability, Deep Learning

1 Introduction

Over the last decade, Machine Learning (ML) - and, by extension, Deep Learning (DL) - has increasingly become prominent in the relationships among people and machines in many Human-Computer Interaction (HCI) areas. NIME research has also aligned with this trend, and several artifacts incorporating or based on neural networks (NNs) have been presented (e.g. [**? ?**]). The NIME community has emphasised the importance of longevity of musical devices, to preserve knowledge [33], promote critical reflections [3], and minimise the environmental impact [35]. In a recent literature review on ML at NIME, Jourdan and Caramiaux have highlighted the main trends and, building upon the reflection on documentation and code sharing by Calegario et al. [4], they argued that "*source code or models is a necessary process to make the system "live", evolve and reuse by the community.*" [30]. In this regard, Deep Generative Models (DGM) represent a peculiar case that highlight the critical importance of comprehensive source code and clear reference materials, as their from-scratch implementation poses unique challenges due to being notoriously data-hungry, requiring high computational resources, and demanding solid practical expertise in the field [**?**].

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

As such, in this paper we further this inquiry conducting a practical investigation into what is required for reuse and replicate DGM-based NIMEs. We actively recreated seven models from the recent literature, either unavailable or partially accessible, relying on the existing resources, code, and documentation. We open-sourced such models, and they are now ready to reuse [1]. By analysing the issues that we encountered in light of the guidelines proposed by Bin [3], we propose a set of documentation best practices. Where systematic reviews make research trends and narratives *legible*, our systematic replication aims to actively reimplement and read *legibility* under the lens of *longevity* and *reuse.*

2 Background

As this work focus on reproducibility of Deep Generative Models, in this section we provide an overview of the literature regarding 1) DL in NIME; 2) longevity and reusability; and 3) documentation as a core component of longevity.

2.1 Deep Learning in NIME

Researchers in music technology have long applied DL models to different tasks, such as classification [19], symbolic generation [23], and audio manipulation and generation [7, 14], and different architectures have been deployed. For instance, Convolutional Neural Networks (CNN) have been used for instrument recognition and data labelling, Long Short Term Memory networks (LSTM) for MIDI sequences generation [12], and DDSP and DDPM for real-time audio generation and data augmentation [10, 28].

In support of such applications, large datasets - such as music21 Bach's chorales[5], Magenta Groove [21] and Maestro [26] or Lakh MIDI [45] - have also been assembled.

In the context of musical interactive systems and NIMEs, NNs has also been used, e.g. for complex mappings [11] or musical improvisation [2]. For instance, Tahiroğlu's AI-terity project exemplifies this paper's focus, initially leveraging a pre-trained GAN model and later retraining it on a custom dataset for its second iteration to better suite the intended artistic context [47]. Recently, Jourdan and Caramiaux [30] present a study analysing 69 Machine/Deep Learning-based contributions from the last 10 years. Notably, the 54% of papers they reviewed do not allow users to adapt data or parameters, and mainly present "black box systems" (ibid); cultural and political implications of the use of ML in NIME are further explored through their companion paper [29]. Drawing from both studies, they propose that

> *"Documentation effort should be accompanied by good practice in ML, and previous work on the*

[1] https://vault.oddworlds.org/s/634PXnGcoQmqStL

> *construction and documentation of datasets and models is a good starting point for the community to build on."* [30]

2.2 Longevity and Reusability

In 2017, Morreale and colleagues highlighted that a significant number of NIMEs are only used in very few performances [39], and investigated factors influencing whether an instrument continues to be used over time. They found that while some instruments persist through adaptation and reuse, others are abandoned due to a range of reasons, including lack of technical support, limited documentation, and incompatibilities with newer technologies. Since then, reflections on NIMEs obsolescence have included sustainable practices [34], longevity as a condition for virtuosity [39], and replication as a learning process [53]. Furthermore, NIME research has explored how instruments remain in use beyond their initial design: repertoire building [17] embeds instruments in musical practice, while hardware durability [37] and opposing "disposable instruments" [6] address sustainability concerns. Strategies for overcoming component obsolescence [48] help extend usability despite shifting technologies, and artistic updates [31] highlight how creative reinterpretation can sustain relevance. More broadly, promoting longevity [42] ensures instruments remain accessible, adaptable, and valuable to future practitioners.

2.3 Documentation and Replicability

Marques-Borbon and Avila observed NIMEs that are "only maintained either anecdotally or as paper citations" [33]. Attempting to remedy this condition, they suggest efforts may be directed towards teaching and performance, where the instruments become activated as performers and students engage with, modify, and adapt them, contributing to extend their development. Similarly, in the context of circuit-bent devices, Dorigatti and Masu suggested applying reuse recursively throughout the design, "favouring future disassemblage and reuse of the components" [9]. In both cases, the authors emphasised the pivotal role of documentation. Calegario and colleagues further raised concerns with the lack of supporting material:

> *"Instruments become available for users in different locations and cultures and future generations, enabling more performers, composers, and audiences to experience artifacts or systems"* [4]

In this regard, Bin [3] proposed five documenting strategies that can help build a critical NIME history: *Collaborative* - promoting collaboration among researchers and practitioners; *Ongoing* - ensuring regular updates and continuous conversation; *Flexible* - adaptable to different contexts and needs; *Openness* - transparency and accessibility; *Complete* - thorough coverage of essential aspects.

While a debate around documentation practices has been developing within the NIME community over the past few years, it has not yet systematically engaged with DL-based NIMEs. On the contrary, outside NIME, several works outlined approaches to document models and datasets (e.g. [18, 38]). These proposals highlight the communication of the characteristics of the model to avoid mismatches between training data and deployment context, rather than documentation for replication purposes.

3 Re-use, Re-train, Re-create

We undertook two sequential processes to advance the understanding of reuse potential and replicability in DGMs presented at NIMEs. Firstly, we scrutinised papers in the relative literature to assess the availability of accompanying resources such as code, data, and documentation. Secondly, for systems that were no longer functional - for example, due to missing or incomplete code - we attempted to reimplement and retrain them from scratch.

3.1 Checking what is available

In our assessment of the available materials necessary for the reuse of DGM components, we first identifed DGMs works presented at NIME. We relied on the aformentioned recent literature review [30], which systematically analysed ML in NIME, and select form the papers they identifies the subset of DGMs. This produced a collection of 17 papers to be scrutinised: [2, 11, 12, 20, 22, 25, 32, 40, 41, 43, 44, 46, 47, 49–52].

For every work, we assessed whether reference to three critical aspects for reuse - **code**, **model weights**, and **datasets** - were available [**?**].

Our analysis revealed that most papers involved training new models, with a smaller portion reusing existing ones. We excluded papers that used private datasets from further consideration: among the remaining ones, making use of public datasets, only a few provided accessible code or model weights, and many links were inactive. As summarized in Figure 1, this left us with seven papers requiring full recreation due to missing or incomplete resources.

We propose three possible ways of reactivate DGMs (or AI in general), ranging from minimal to maximal effort required:

- **Re-use**: using existing code and pretrained models;
- **Re-train**: using existing code and data;
- **Re-create**: implement architectures from scratch.

3.1.1 Re-use: Code & Weights. The first possible reactivation is using the system without needing any additional work - in line with what recently proposed on NIME longevity [35] - relying on the availability of both code and model weights. While the code usually defines the model architecture and data processing steps, the weights (typically a checkpoint file) contain the final state of the trained model, required to replicate the learned behaviour.

3.1.2 Re-train: Code & Data. A second approach to reactivation involves retraining models. Although gaps in documentation (e.g. missing hyperparameters or the use of deprecated libraries) can lead to significant deviations from the intended outcomes, these are usually edge cases: when both the code and dataset are available, retraining often offers a feasible path to reactivation, typically requiring moderate effort and technical familiarity.

3.1.3 Re-create: Data. Whether the code is not available and only the training data accessible, the technical details and functionalities need to be inferred from supplementary resources, primarly, the original publication itself, or residual code blocks from related works or repositories. This approach is similar to the one derived from museology and media study for NIME preservation, which highlight the importance of integrating together multiple resources to recreate a system's functionality and context [**?**].

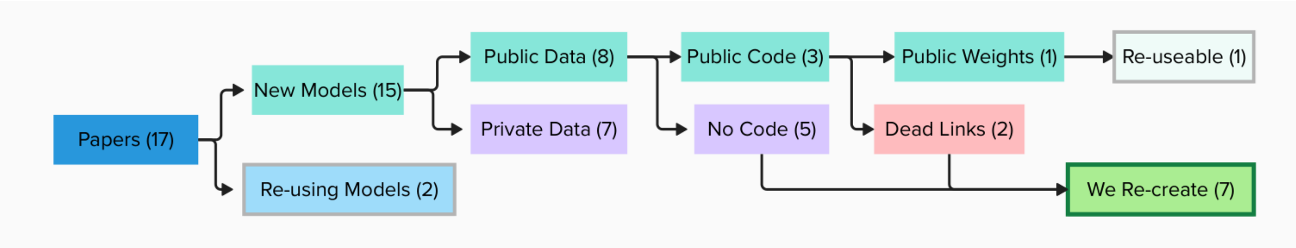

Figure 1: Breakdown of our process choosing the papers to re-create.

3.2 Re-create

In this paper, we concentrate on the last approach - re-create - as a way to practically investigate to which extent a paper publications can be sufficient to reactivate a system. We acknowledge that the contribution of a paper lies beyond the availability of the code, and we do not aim to propose a critique to previous publications. We simply use those paper that do not have the related code available as a case study to investigate longevity and documentation of DGMs. Indeed, while we advocate for reusability, we acknowledge it may not always align with a paper's core contribution.

Thus, we decided to recreate systems that could be replicated based on the available material and descriptions. As mentioned in Sec. 3.1, out of the papers selection, 7 did not provide a link to a working code but still providing access to the data, thus fitting in the condition for re-creation [2, 12, 20, 32, 43, 49, 50]. We recreated these models in PyTorch [2], due to its flexibility, ease of implementation, and widespread adoption in the DL community.

Notably, all the models considered were based on sequence generation of symbolic music. As such, as a general principle, we employed a modular approach, allowing components to be reused in different combinations and with different parameters. The 7 papers relied on data sources that include - but not limited to - a series of well-established MIDI datasets: Magenta Groove [21], Magenta Maestro [26], music21 Bach Chorales [5], and Lakh MIDI [45].

To recreate these models, we followed a systematic implementation approach. Over the course of one month, we methodologically approached each paper we considered: first, we read and analysed the text, isolating unfamiliar or unclear aspects in the architecture and cross-referencing similar ones on GitHub; then, we developed the code, initially constituted of simple experimental scripts trained locally. The goal was not to fully optimise the models, rather check for viability and identify errors or inconsistencies. In the case of missing details, we attempted a reasonable approximation based on other resources. then id take a best guess from what i can read elsewhere. Also, when data or tokenisation strategies were not available, we promoted consistency by readapting existing preprocessing pipelines.

While broadly sharing a common generative goal, such models still vary in their underlying architectures and implementation details. An in-depth description of the pipelines involved is beyond the scope of this work: for technical details, we invite the reader to refer to our code and to the respective original contributions. However, for the sake of clarity, we still provide a comprehensive summary of the main the technical aspects of the models considered - Table 1.

[2]https://pytorch.org/

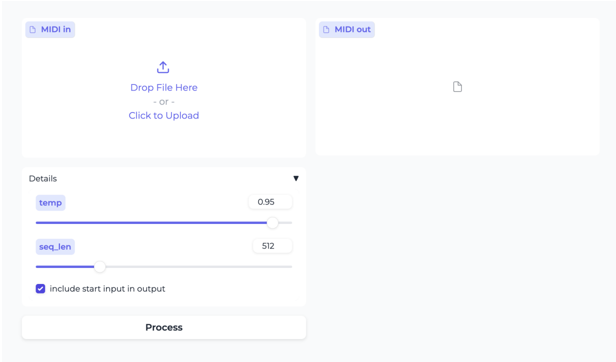

Figure 2: Screenshot of web interface used for testing models

3.2.1 *Web Interface.*
Alongside our implementations of the models, we created Gradio [1] interfaces for each of the models - Fig. 2. These interfaces allow to drag-and-drop midi files for transformation, generate samples, and adjust model parameters. It is important to note these interfaces are distinct from the original authors designs and do not represent the intended interaction: instead, these serve as a utility to streamline and normalise the inference process.

4 RE-create: Critical Aspects

Drawing from the recreation process we undertook, in this section we outline the main aspects and challenges to be considered for DGMs reactivation.

Data Preparation Challenges

Well-curated datasets are as key a contribution as the novelties of model architecture. Indeed, it has been shown how different models yield similar benchmarks when trained on the same data [24], indicating that complex models don't necessarily offer improvements over simpler ones. Also, Warren and Çamcı discussed style loss and diversity in relation to choice of dataframes and data processing [50].

Data preparation involves multiple stages from to filter and refine to restructuring the data, e.g. augmentations, transposition, or sequence padding. In sequence models like the ones we focused on, tokenisation is often a key element. Each paper used particular tokenisation algorithms to represent the MIDI data as a sequence of numbers. In our process, where possible, we simplify our pipeline using standard libraries - e.g. *miditok* [15].

Filtering methods, such as restricting datasets by key or timing, can also increase complexity. A key issue was unknown sequence lengths — whether to pad or chunk, often overlooked in the provided documentation. As a result, our training sets may

Paper Ref	Year	Architecture	Parameters	Input Type	Output Type	Key Parameters
[49]	2017	RBM	32,564	64D binary vector	64D binary vector	Visible units: 64 Hidden units: 500
[12]	2019	Seq2Seq BiLSTM	1,785,026	Sequential tokens	Sequential tokens	Vocab size: 130 Hidden dim: 256 Embed dim: 32
[2]	2020	Dual LSTM	4,804,909	Rhythm + CPC + MIDI tokens	Token + Key prediction	Hidden dim: 400 Embed dim: 50 Token dim: 135
[20]	2021	Dual LSTM-VAE	7,968,822	Score (9D) + Rhythm (18D)	Joint 27D sequence	Hidden dim: 512 Latent dim: 256
[43]	2021	Transformer-XL	18,944,296	Token sequence	Next token prediction	Model dim: 512 Depth: 6 Heads: 8
[32]	2021	Transformer	48,055,312	Token sequence	Next token prediction	Model dim: 512 Depth: 12 Heads: 8 Vocab size: 10000
[50]	2022	ConvVAE	3,148	100x100 binary image	5x16 velocity matrix + 8D Markov state	Latent dim: 8 3 conv layers 2 decoders

Table 1: A summary of the technical details of papers we selected for recreation.

differ from the originals, preventing exact replication. Overall, we support that ML papers should indeed prioritise clarity in defining datasets and data preparation methods.

Implementation Complexities

We often encountered lack of precise descriptions, from large-scale architectural details - e.g. number of layers or types of activation functions - to hyper-parameters - e.g. regularisation strategies or learning rates. Inconsistencies were also common with regard to optimisers and loss functions. We acknowledge that addressing all of the required information in any NIME paper may be inappropriate and irrelevant to the main contributions of that specific work: this reinforces the benefit of supplementary files.

Computational Resources

In the recreation of large models (as the generative ones usually are), one of the main issues is related to the hardware infrastructure (dedicated GPUs). The limitations of hardware has been indeed acknowledged in the papers we replicated, with DGM models being inhibited by the "expense of computational resources needed to train them" [20]. While the use of a non-CUDA environment knowingly restricts development and does not often represent a viable solution [**?**], it may be worth considering in what may be expected of a general audience attempting to rework the models with varying constraints.

Metrics for Models Assessment

Assessing recreated models is challenging due to the lack of standard metrics; as noted by Jourdan and Caramiaux [30], evaluation often relies on subjective listening tests (e.g. [12, 32, 43, 49]). While these evaluations fruitfully capture artistic and experiential qualities, their lack of objectivity may limit comparability and reproducibility, especially in generative systems where outputs are inherently variable and results often inferred via demonstrations or author commentary.

Establishing clear approaches to evaluation, possibly combining objective metrics with subjective insights, could support more reliable assessment and reuse of DGMs in future NIME work.

Impact of Framework Choice

An unclear influence on our results is the choice of programming languages and frameworks. The original papers used various DL libraries; to ensure compatibility and promote future reuse, we harmonised our work within the PyTorch framework. While this choice helped streamline our process, it may have subtly shaped our working habits and implementation logic. Rather than recreating each system's original interface, we chose to normalise interactions across models, which led to necessary algorithmic adjustments - e.g. in sampling and generation functions. The role of technology-related choices is broadly noted in several papers we reviewed. For example, Erdem et al. state: "Technology-related choices, such as choosing a particular sensor or algorithm, inevitably become compositional choices." [11]. How these choices affect the models and inference pipelines, and how they intertwine with availability and reuse, is hard to quantify. Finally, algorithmic inscriptions and uncertainties may also shape the outputs of our models, as in sound design choices [36].

Authors' Knowledge

Finally, we reflect on the knowledge assumed by the authors of the original papers, their expectations towards the reader's expertise, and our own as re-creators. In keeping up with rapid DGM developments, it is unclear what level of technical familiarity is expected from the audience and what constitutes common knowledge today may have been novel or inaccessible even a year ago. With the range of cross-disciplinary work presented in NIME, addressing technical detail appropriately can be difficult and could be a barrier to understanding and replicating the work, especially for those new to the field or working from a different disciplinary background. While this is probably intrinsic related to the rapid DGM development we are facing, we argue that how we communicate technology remains an important point that need to be addressed by the NIME community, contributing to preserving a deep intimate collaboration between music and technology.

5 Implication and Practices

Building up on the insights we gathered from our reactivation process, we identified recurrent issues when engaging with DGMs. In the past years, many NIME scholars have argued in favour of longevous devices [33], properly documented [4] and used over time to foster knowledge generation emerging from prolonged entangled interactions (e.g. [13, 31]). Reflecting on the difficulties of replicating DGM systems has prompted us to suggest some guidelines to foster their longevity in NIME. The main aim of these guidelines is to encourage greater availability and reuse of individual components, allowing for more collaborative experimentation and wider understanding.

We acknowledge that these design implications are developed in reference to older works, and we are pleased that more recently the adoption of similar ideas to the ones we present can be seen across the ML community. Sharing of code and model weights has become more common today, the required infrastructure for sharing is far more available, and commenting on and adapting systems ever more active.

Overall, we propose four guidelines for preserving DGM NIMEs, which resonates with Bin's five suggestions [3] for a critical history of NIME - Fig. 3.

Availability of Data (Collaborative, Openness)
Using datasets such as Magenta or music21 can provide a useful common foundation for models development. In contrast to this, the personalisation of models is attractive to creative practitioners, and choosing to use small personal datasets can "avoid the problem of normativity and external biases that would stem from big data" [29]. In cases where the whole training dataset cannot be made available, we would encourage providing the data collection along with preparation methods and algorithms. Similarly, adoption of augmentation tools or tokenisation libraries, such as miditok [15] allow for less repetitive work in setting up data processing pipelines.

Architectures and HyperParams (Complete, Flexible)
Comprehensive documentation of training parameters, such as sequence lengths, batch sizes, or memory usage, is key information in replicating the performance of a model, as small adjustments to these can cause large changes in outputs. Making this information available in supplementary materials would allow future adaptations to replicate training setups, and better understand the effect of changes to these parameters.

Availability of Weights (Collaborative, Openness)
Sharing pre-trained model weights offer clear ecological and economic benefits, while reducing time needed in replicating or adapting DGM NIMEs. The availability of pre-trained weights opens up new paths for exploration, enabling researchers to maintain focus on experimentation with interfaces or exploration of the model's potential, finding novel features or applications, "allowing the user to begin creating quickly without any time-consuming training required". [32]

Design for Reuse (Flexible, Ongoing)
In designing DGM NIMEs it is important to consider the future reuse of the models. By anticipating future users and uses since the early development stages, we can help create tools and systems that not only serve our immediate intents, but can open up possibilities for others: "You may not be able to design for the unexpected, but you can design to allow the unexpected" [8]. As such, each contribution can serve as a stepping stone for others, enabling reassembly and reconfiguration to learn how these NIMEs function and imagine how they can evolve.

We would not want to prescribe specific standards, but encourage consideration of these general themes. Previous discussion of documenting and sharing code alongside papers raised the issue that "Deferring to web links for this material isn't an adequate solution because it disassociates writing exactly from that material we suggest can benefit most from peer review for practitioners" [16]. It seems important to integrate these resources within the academic framework to support their persistence and accessibility, as we have seen during our research many web-links that have expired or are missing materials. Yet the practicality of this integration remains a challenge: overall, we would encourage further experimentation with documentation and availability of models to find appropriate ways of sharing and assessing this material.

6 Future Directions

Building on our experience we propose several potential paths for future explorations.

Exploring new interfaces - Our work unified these models under one type of interaction and inference. Future explorations could involve designing new interfaces that reshape these models, adjusting how input features and conditions influence behavior. As noted, "variables can be exposed in different ways within user interfaces" [20], which can guide models towards preferred outputs.

Retraining on different data - As we discussed in Subsec. 4, authors have expressed concerns on cultural scopes and dataset limitations. Since "the choice of conditioning variables (along with the choice of training data) outlines an initial set of limitations that define how a model might be used" [20], retraining and fine tuning models with different datasets could enhance model versatility and relevance to other musical traditions.

Investigation into models - Delving deeper into the models - trough features exploration, latent manipulation, or creative (mis)use - could uncover new possibilities. As Jourdan and Caramiaux suggest, "going beyond the framework of systematic review to open the analysis to other archives", recreating these models enables reassessment and further discovery [30].

Frankenime - Finally, our recreation process leads us to the idea of *Frankenime*, a speculative concept emerging from the possible recombination and repurpose of models and datasets. While still at an early stage, we draw from the idea of "zombie-media" [27] that illustrates the unintended outcomes of modular reuse. This resonates with the work by Murrary-Browne and Tigas, who noted: "when encountering novelty through opaque processes such as ML, it can be unclear whether what has been created is a readily identifiable and imitable artefact of the tool (such as a synth preset) or a unique avenue worth staking our creative identity upon" [40]. Building on this, we question if these deranged qualities emerge from the tools themselves, our missteps in re-creation, or as artifacts of a new composition coming

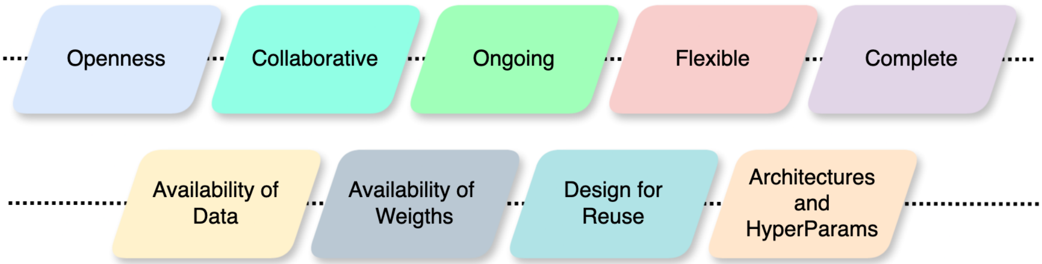

Figure 3: Bin's five suggestions for a critical history of NIME [3] over our four suggestions for preserving DGM NIMEs.

to life? Does recursive reuse create disfigured Frankenimes or graceful resurrections? Whether this is an indication of paths to be pursued remains an open question.

6.1 Conclusion

In this paper, we first evaluated whether DGM systems presented at NIME are still available, and afterward recreated a subset of the non-available models. We have observed a need for more documentation and supplementary material of DGM NIMEs, and explored the consequences of the availability of documentation through the replication and retraining of seven models, which we open-sourced and made available for the community. Additionally, by reflecting on our recreation process in light of reflection on critical discourse on documentation in NIME, we developed a number of strategies for NIME DGM research documentation. We hope that this can lead to new critical and longevous ML research in NIME.

Reducing the need for people to replicate the training process is of both economic and environmental concern [**?**]. Indeed, the training process is often the most energy-intensive part of DGMs. Recently, we can observe more active sharing of models with the popularity of services like huggingface[3]. With this work we hope to contribute to the creation of similar sharing platforms for NIME.

7 Ethical Standards

In this research, we chose to recreate only the works that made use of publicly available datasets, excluding systems that used custom or private data of the original authors. As the performance of these models was of less interest to us than the process of recreating, we chose to work only on our laptop to minimise energy usage in training models. Our aim was not to recreate the works entirely, but investigate challenges in recreating small parts of DGMs that sit within greater contributions. Albeit we did not contacted the original authors during our recreation process, relying solely on materials that were publicly accessible at the time of writing, we still recognise the value of such communication and welcome dialogue on responsible reuse and authorship within the community. Our thanks go to the original authors for their great work.

References

[1] Abubakar Abid, Ali Abdalla, Ali Abid, Dawood Khan, Abdulrahman Alfozan, and James Zou. 2019. Gradio: Hassle-Free Sharing and Testing of ML Models in the Wild. *arXiv preprint arXiv:1906.02569* (2019).

[2] 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*, Romain Michon and Franziska Schroeder (Eds.). Birmingham City University, Birmingham, UK, 635–640. https://doi.org/10.5281/zenodo.4813234

[3] S. M. Astrid Bin. 2021. Discourse is critical: Towards a collaborative NIME history. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Shanghai, China, Article 8. https://doi.org/10.21428/92fbeb44.ac5d43e1

[4] Filipe Calegario, João Tragtenberg, Christian Frisson, Eduardo Meneses, Joseph Malloch, Vincent Cusson, and Marcelo M. Wanderley. 2021. Documentation and Replicability in the NIME Community. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Shanghai, China, Article 4. https://doi.org/10.21428/92fbeb44.dc50e34d

[5] Michael Scott Cuthbert and Christopher Ariza. 2010. Music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data.. In *ISMIR*, J. Stephen Downie and Remco C. Veltkamp (Eds.). International Society for Music Information Retrieval, 637–642. http://dblp.uni-trier.de/db/conf/ismir/ismir2010.html#CuthbertA10

[6] Tom Davis. 2017. The Feral Cello: A Philosophically Informed Approach to an Actuated Instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Aalborg University Copenhagen, Copenhagen, Denmark, 279–282. https://doi.org/10.5281/zenodo.1176250

[7] Prafulla Dhariwal, Heewoo Jun, Christine Payne, Jong Wook Kim, Alec Radford, and Ilya Sutskever. 2020. Jukebox: A Generative Model for Music. arXiv:2005.00341 [eess.AS]

[8] Alan Dix. 2007. Designing for Appropriation. In *Proceedings of BCS HCI 2007, People and Computers XXI*, Vol. 2. BCS eWiC. https://www.alandix.com/academic/papers/papers/HCI2007-appropriation/

[9] Enrico Dorigatti and Raul Masu. 2022. Circuit Bending and Environmental Sustainability: Current Situation and Steps Forward. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. The University of Auckland, New Zealand, Article 5. https://doi.org/10.21428/92fbeb44.18502d1d

[10] Jesse Engel, Lamtharn (Hanoi) Hantrakul, Chenjie Gu, and Adam Roberts. 2020. DDSP: Differentiable Digital Signal Processing. In *International Conference on Learning Representations*. https://openreview.net/forum?id=B1x1ma4tDr

[11] Cagri Erdem, Benedikte Wallace, and Alexander Refsum Jensenius. 2022. CAVI: A Coadaptive Audiovisual Instrument–Composition. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. The University of Auckland, New Zealand, Article 25. https://doi.org/10.21428/92fbeb44.803c24dd

[12] Andrei Faitas, Synne Engdahl Baumann, Torgrim Rudland Næss, Jim Torresen, and Charles Patrick Martin. 2019. Generating Convincing Harmony Parts with Simple Long Short-Term Memory Networks. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Marcelo Queiroz and Anna Xambó Sedó (Eds.). UFRGS, Porto Alegre, Brazil, 325–330. https://doi.org/10.5281/zenodo.3672980

[13] Rebecca Fiebrink and Laetitia Sonami. 2020. Reflections on Eight Years of Instrument Creation with Machine Learning. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Romain Michon and Franziska Schroeder (Eds.). Birmingham City University, Birmingham, UK, 237–242. https://doi.org/10.5281/zenodo.4813334

[14] Seth* Forsgren and Hayk* Martiros. 2022. Riffusion - Stable diffusion for real-time music generation. (2022). https://riffusion.com/about

[15] Nathan Fradet, Jean-Pierre Briot, Fabien Chhel, Amal El Fallah Seghrouchni, and Nicolas Gutowski. 2021. MidiTok: A Python package for MIDI file tokenization. In *Extended Abstracts for the Late-Breaking Demo Session of the 22nd International Society for Music Information Retrieval Conference*. https://archives.ismir.net/ismir2021/latebreaking/000005.pdf

[16] Adrian Freed, John MacCallum, and Sam Mansfield. 2013. "Old" is the New "New": a Fingerboard Case Study in Recrudescence as a NIME Development Strategy. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Graduate School of Culture Technology, KAIST, Daejeon, Republic of Korea, 441–445. https://doi.org/10.5281/zenodo.1178524

[17] Takuto Fukuda, Eduardo Meneses, Travis West, and Marcelo M. Wanderley. 2021. The T-Stick Music Creation Project: An approach to building a creative community around a DMI. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Shanghai, China, Article 13. https:

[3]https://huggingface.co

//doi.org/10.21428/92fbeb44.26f33210
[18] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. Datasheets for datasets. *Commun. ACM* 64, 12 (nov 2021), 86–92. https://doi.org/10.1145/3458723
[19] Nicholas Gillian and Joseph A. Paradiso. 2012. Digito: A Fine-Grain Gesturally Controlled Virtual Musical Instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. University of Michigan, Ann Arbor, Michigan. https://doi.org/10.5281/zenodo.1178263
[20] Jon Gillick and David Bamman. 2021. What to Play and How to Play it: Guiding Generative Music Models with Multiple Demonstrations. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Shanghai, China, Article 6. https://doi.org/10.21428/92fbeb44.06e2d5f4
[21] Jon Gillick, Adam Roberts, Jesse Engel, Douglas Eck, and David Bamman. 2019. Learning to Groove with Inverse Sequence Transformations. In *International Conference on Machine Learning (ICML)*.
[22] Jeff Gregorio and Youngmoo E. Kim. 2021. Evaluation of Timbre-Based Control of a Parametric Synthesizer. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Shanghai, China, Article 83. https://doi.org/10.21428/92fbeb44.31419bf9
[23] Gaëtan Hadjeres, François Pachet, and Frank Nielsen. 2017. DeepBach: a steerable model for bach chorales generation. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70* (Sydney, NSW, Australia) *(ICML'17)*. JMLR.org, 1362–1371.
[24] Alon Halevy, Peter Norvig, and Fernando Pereira. 2009. The Unreasonable Effectiveness of Data. *IEEE Intelligent Systems* 24, 2 (2009), 8–12. https://doi.org/10.1109/MIS.2009.36
[25] Lamtharn Hantrakul. 2018. GestureRNN: A neural gesture system for the Roli Lightpad Block. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Thomas Martin Luke Dahl, Douglas Bowman (Ed.). Virginia Tech, Blacksburg, Virginia, USA, 132–137. https://doi.org/10.5281/zenodo.1302703
[26] Curtis Hawthorne, Andriy Stasyuk, Adam Roberts, Ian Simon, Cheng-Zhi Anna Huang, Sander Dieleman, Erich Elsen, Jesse Engel, and Douglas Eck. 2019. Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset. In *International Conference on Learning Representations*. https://openreview.net/forum?id=r1lYRjC9F7
[27] Garnet Hertz and Jussi Parikka. 2012. Zombie Media: Circuit Bending Media Archaeology into an Art Method. *Leonardo* 45, 5 (10 2012), 424–430. https://doi.org/10.1162/LEON_a_00438 arXiv:https://direct.mit.edu/leon/article-pdf/45/5/424/1575092/leon_a_00438.pdf
[28] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 6840–6851. https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf
[29] Théo Jourdan and Baptiste Caramiaux. 2023. Culture and Politics of Machine Learning in NIME: A Preliminary Qualitative Inquiry. , Article 47 (May 2023). http://nime.org/proceedings/2023/nime2023_47.pdf
[30] Théo Jourdan and Baptiste Caramiaux. 2023. Machine Learning for Musical Expression: A Systematic Literature Review. , Article 46 (May 2023). http://nime.org/proceedings/2023/nime2023_46.pdf
[31] Hans Leeuw. 2021. Virtuoso mapping for the Electrumpet, a hyperinstrument strategy. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Shanghai, China, Article 39. https://doi.org/10.21428/92fbeb44.a8e0cceb
[32] Jeffrey A. T. Lupker. 2021. Score-Transformer: A Deep Learning Aid for Music Composition. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Shanghai, China, Article 59. https://doi.org/10.21428/92fbeb44.21d4fd1f
[33] Adnan Marquez-Borbon and Juan Pablo Martinez-Avila. 2018. The Problem of DMI Adoption and Longevity: Envisioning a NIME Performance Pedagogy. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Thomas Martin Luke Dahl, Douglas Bowman (Ed.). Virginia Tech, Blacksburg, Virginia, USA, 190–195. https://doi.org/10.5281/zenodo.1302541
[34] Raul Masu, Adam Pultz Melbye, John Sullivan, and Alexander Refsum Jensenius. 2021. NIME and the Environment: Toward a More Sustainable NIME Practice. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Shanghai, China, Article 24. https://doi.org/10.21428/92fbeb44.5725ad8f
[35] Raul Masu, Fabio Morreale, and Alexander Refsum Jensenius. 2023. The O in NIME: Reflecting on the Importance of Reusing and Repurposing Old Musical Instruments. , Article 14 (May 2023). http://nime.org/proceedings/2023/nime2023_14.pdf
[36] Andrew McPherson and Giacomo Lepri. 2020. Beholden to our tools: negotiating with technology while sketching digital instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Romain Michon and Franziska Schroeder (Eds.). Birmingham City University, Birmingham, UK, 434–439. https://doi.org/10.5281/zenodo.4813461
[37] Nicolo Merendino, Giacomo Lepri, Antonio Rodà, and Raul Masu. 2023. Redesigning the Chowndolo: a Reflection-on-action Analysis to Identify Sustainable Strategies for NIMEs Design. , Article 27 (May 2023). http://nime.org/proceedings/2023/nime2023_27.pdf
[38] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) *(FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 220–229. https://doi.org/10.1145/3287560.3287596
[39] Fabio Morreale and Andrew McPherson. 2017. Design for Longevity: Ongoing Use of Instruments from NIME 2010-14. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Aalborg University Copenhagen, Copenhagen, Denmark, 192–197. https://doi.org/10.5281/zenodo.1176218
[40] Tim Murray-Browne and Panagiotis Tigas. 2021. Latent Mappings: Generating Open-Ended Expressive Mappings Using Variational Autoencoders. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Shanghai, China, Article 66. https://doi.org/10.21428/92fbeb44.9d4bcd4b
[41] Torgrim Rudland Næss and Charles Patrick Martin. 2019. A Physical Intelligent Instrument using Recurrent Neural Networks. In *19th International Conference on New Interfaces for Musical Expression, NIME 2019, Porto Alegre, Brazil, June 3-6, 2019*. nime.org, 79–82. http://www.nime.org/proceedings/2019/nime2019_paper016.pdf
[42] Alex Nieva, Johnty Wang, Joseph Malloch, and Marcelo Wanderley. 2018. The T-Stick: Maintaining a 12 year-old Digital Musical Instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Thomas Martin Luke Dahl, Douglas Bowman (Ed.). Virginia Tech, Blacksburg, Virginia, USA, 198–199. https://doi.org/10.5281/zenodo.1302545
[43] Thomas Nuttall, Behzad Haki, and Sergi Jorda. 2021. Transformer Neural Networks for Automated Rhythm Generation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Shanghai, China, Article 33. https://doi.org/10.21428/92fbeb44.fe9a0d82
[44] Rohan Proctor and Charles Patrick Martin. 2020. A Laptop Ensemble Performance System using Recurrent Neural Networks. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Romain Michon and Franziska Schroeder (Eds.). Birmingham City University, Birmingham, UK, 43–48. https://doi.org/10.5281/zenodo.4813481
[45] Colin Raffel. 2016. *Learning-Based Methods for Comparing Sequences, with Applications to Audio-to-MIDI Alignment and Matching*. PhD dissertation. Columbia University.
[46] Koray Tahiroğlu, Miranda Kastemaa, and Oskar Koli. 2020. Al-terity: Non-Rigid Musical Instrument with Artificial Intelligence Applied to Real-Time Audio Synthesis. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Romain Michon and Franziska Schroeder (Eds.). Birmingham City University, Birmingham, UK, 337–342. https://doi.org/10.5281/zenodo.4813402
[47] Koray Tahiroğlu, Miranda Kastemaa, and Oskar Koli. 2021. AI-terity 2.0: An Autonomous NIME Featuring GANSpaceSynth Deep Learning Model. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Shanghai, China, Article 80. https://doi.org/10.21428/92fbeb44.3d0e9e12
[48] Sam Trolland, Alon Ilsar, Ciaran Frame, Jon McCormack, and Elliott Wilson. 2022. AirSticks 2.0: Instrument Design for Expressive Gestural Interaction. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. The University of Auckland, New Zealand, Article 51. https://doi.org/10.21428/92fbeb44.c400bdc2
[49] Richard Vogl and Peter Knees. 2017. An Intelligent Drum Machine for Electronic Dance Music Production and Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Aalborg University Copenhagen, Copenhagen, Denmark, 251–256. https://doi.org/10.5281/zenodo.1176238
[50] Nick Warren and Anil Çamci. 2022. Latent Drummer: A New Abstraction for Modular Sequencers. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. The University of Auckland, New Zealand, Article 38. https://doi.org/10.21428/92fbeb44.ed873363
[51] Aline Weber, Lucas Nunes Alegre, Jim Torresen, and Bruno C. da Silva. 2019. Parameterized Melody Generation with Autoencoders and Temporally-Consistent Noise. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Marcelo Queiroz and Anna Xambó Sedó (Eds.). UFRGS, Porto Alegre, Brazil, 174–179. https://doi.org/10.5281/zenodo.3672914
[52] Ruihan Yang, Tianyao Chen, Yiyi Zhang, and gus xia. 2019. Inspecting and Interacting with Meaningful Music Representations using VAE. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Marcelo Queiroz and Anna Xambó Sedó (Eds.). UFRGS, Porto Alegre, Brazil, 307–312. https://doi.org/10.5281/zenodo.3672974
[53] Luis Zayas-Garin, Jacob Harrison, Robert Jack, and Andrew McPherson. 2021. DMI Apprenticeship: Sharing and Replicating Musical Artefacts. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Shanghai, China, Article 26. https://doi.org/10.21428/92fbeb44.87f1d63e