CHAPTER

Explainable AI and Music

N. Bryan-Kinns[a], B. Banar[b], C. Ford[b], C. N. Reed[c], Y. Zhang[b], J. Armitage[d]

[a]University of the Arts London, UK; [b]Loughborough University London, UK; [c]King's College London, UK; [d]Iceland University for the Arts, Iceland

ARTICLE HISTORY
Compiled August 8, 2024; Published August 29, 2024

Bryan-Kinns, N., Banar, B., Ford, C., Reed, C. N., Zhang, Y., Armitage, J. (2024). Explainable AI and Music. In Mou, L. (Ed.), *Artificial Intelligence for Art Creation and Understanding* (pp. 1-29). CRC Press. DOI: 10.1201/9781003406273-1

ABSTRACT
The field of eXplainable Artificial Intelligence (XAI) has become a hot topic examining how machine learning models such as neural nets and deep learning techniques can be made more understandable to humans. However, there is very little research on XAI for the arts. This chapter explores what XAI might mean for AI and art creation by exploring the potential of XAI for music generation. One hundred AI and music papers are reviewed to illustrate how AI models are being explained, or more often not explained, and to suggest some ways in which we might design XAI systems to better help humans to get an understanding of what an AI model is doing when it generates music. Then the chapter demonstrates how a latent space model for music generation can be made more explainable by extending the MeasureVAE architecture to include explainable attributes in combination with offering real-time music generation. The chapter concludes with four key challenges for XAI for music and the arts more generally: i) the nature of explanation; ii) the effect of AI models, features, and training sets on explanation; iii) user centred design of XAI; iv) Interaction Design of explainable interfaces.

KEYWORDS
Explainable AI; Human-Centred AI; Generative AI; AI for music

With few exceptions, music has been for some centuries the art which has devoted itself not to the reproduction of natural phenomena, but rather to the expression of the artist's soul, in musical sound.

(Kandinsky, 1912)

CONTACT N. Bryan-Kinns Email: nickbk@acm.org

1. Introduction

Music is a fundamental element of intangible cultural heritage and a form of artistic expression found in all cultures. It has been an application of Artificial Intelligence (AI) since before the birth of digital computers with the potential of computing machines to create complex music noted as early as 1843 by Ada Lovelace (1843). AI is now deployed across all the culture cycle stages (Pessoa & Deloumeaux, 2009) of music creation, production, dissemination, exhibition/ reception/ transmission, and consumption/ participation. In this chapter we focus on AI for music creation (generation) rather than its role in other culture cycle stages such automated mixing (De Man, Reiss, & Stables, 2017) in music production.

Early uses of AI for music creation typically employed probabilistic models to compose music offline, e.g. the Illiac Suite composed by AI created by Lejaren Hiller and Leonard Isaacson in 1957 (Hiller & Isaacson, 1957). Increased computing power and speed over the following decades opened up opportunities for real-time music generation by AI and human-computer interaction with probabilistic AI models (Pachet, 2003). Machine Learning, and more specifically Deep Learning models have gained popularity in recent years as larger datasets and more computational power have become increasingly available. Although Deep Learning generative AI models have been shown to produce compelling music (e.g. Dinculescu, Engel, and Roberts (2019)'s MidiMe), they require large amounts of training data (usually more than 1m musical notes) and computational time (often days) and are typically difficult to control in meaningful ways (Bryan-Kinns et al., 2021). In contrast, probabilistic approaches such as multi-order multi-feature Markov models e.g. Whorley and Laney (2020) offer more controllable music generation with lower dataset and computational requirements offering more opportunity for real-time interaction, but their outputs are often less novel.

Regardless of the underlying AI paradigm, the complex nature of generative AI models typically require users have in-depth knowledge of AI techniques in order to successfully manipulate the models to create new content beyond trial and error. Compounding this problem, generative music models typically do not offer user interfaces which provide insight into how the music is generated or how the generation can be controlled or influenced. The result of this is that many generative models leave musicians feeling disconnected from the creative process, making them inaccessible to anyone besides the creator, much as is found with Digital Musical Instruments (Benyon & Macaulay, 2002; Morreale & McPherson, 2017; Wallis, Ingalls, Campana, & Vuong, 2013). Without an understanding of their input or succinct ways to influence models, artists may also be unable to find or own their own role in the creative process — whether as collaborator, controller, designer, or merely recipients of generated output (Long, Padiyath, Teachey, & Magerko, 2021a; Louie, Coenen, Huang, Terry, & Cai, 2020b).

In this chapter we explore how generative music models can be made more understandable and controllable by users, especially users who are musicians. We focus on the *use* of AI rather than creation of new AI models per se and engage directly with discourse on Human-Centred AI (HCAI; (Garibay et al., 2023)), specifically the concerns of explainable AI (XAI (Gunning, 2016)) and how AI models can be created which are more interpretable by humans. First we outline XAI and HCAI research and their potential to inform more human-centred generative AI music systems. We then summarise a systematic review of the explainability of 100 generative AI models for music to illustrate state-of-the-art. Following this review we introduce our implementation of an XAI generative music system and use this to illustrate the potential of XAI for music. We conclude by identifying key research challenges in XAI for music generation, and XAI for the arts more broadly.

2. Related Work

Music is made and appreciated by humans and can be generated by AI. Whilst approaches such as Computational Creativity (Colton & Wiggins, 2012) explore the use of AI to autonomously generate music (Carnovalini & Rodà, 2020) there has been notably less research on how AI is designed and used for music making. This manifests itself in an emphasis on AI autonomy and agency with little or no consideration for human agency and value in the fundamentally human and socially constructed endeavour of music making. For example, there is a lack of research on how generative AI is used in music making practice (Xambó, 2022) or as new forms of musical instruments (Pelinski et al., 2022), and a lack of research on how human and AI agency might contribute co-creation of music. To begin to address this gap, Jourdan and Caramiaux (2023a, 2023b) recently published initial surveys of AI/ ML in the New Interfaces for Musical Expression (NIME) community, kicking off an important process of exploring the role of AI in music. A promising route to redress this imbalance is to draw on research from Human-Centered AI (HCAI; Garibay et al. (2023); Shneiderman (2022)) given the inherently human-centered nature of music and music making. HCAI aims to ensure that "advances in AI augment rather than replace humans and improve their environment" (Garibay et al., 2023), with the goal to "amplify, rather than erode, human agency" (Shneiderman, 2022). However, to date HCAI research has predominantly focused on task-oriented AI applications such as medical decision making or self-driving cars, and there is little research on HCAI for more open ended activities such as the arts or creative practice. In this chapter we explore the potential for HCAI approaches to make generative AI more human-centred. In particular we explore one of the grand challenges of HCAI (Garibay et al., 2023) - how to design more human-centered Human-AI Interaction.

A key issue for designing Human-AI Interaction is the increasing complexity and obtuseness of AI models. This is a challenge which Explainable AI (XAI) (Gunning, 2016; Kozierok et al., 2021) sets out to tackle by making complex and difficult to understand AI models such as neural networks and deep learning models more understandable to humans. In our view, increasing the explainability of a generative music model supports increased human agency in music making with AI through, for example, increased sense of control and influence over the AI music generation, and increased levels of understanding and engagement with the music output and the AI itself. In this view AI is framed as a semi-autonomous generative agent/ tool/ instrument for co-creating music rather than an opaque and autonomous music generating system. XAI approaches include generating understandable explanations of AI model behaviour, structuring and labelling features of AI models to make them more understandable, and creating simpler, more understandable models which approximate the behaviour of complex AI models (ibid.). However, as with HCAI more broadly (Garibay et al., 2023), XAI research has predominantly examined the explainability of task-oriented AI systems such as AI medical diagnosis (Quellec et al., 2021), autonomous vehicles (Du et al., 2019; Shen et al., 2020), and AI models of consumer behaviour (Sajja et al., 2021) with no research exploring XAI for the arts and creativity to date.

Unfortunately the term *explanation* is ambiguous and variously defined in XAI research, but nevertheless a brief review of a few noteworthy definitions helps to contextualise the idea as used in this chapter. Ciatto, Schumacher, Omicini, and Calvaresi (2020) defines *explaining* as being "the activity of producing a more interpretable object X' out of a less interpretable one, namely X" where *interpreting* X is the activity of "assigning a subjective meaning to X" (ibid.). This aligns with Guidotti et al. (2018)'s definition of the *interpretability* of an AI model as the ability "to provide the meaning in understandable terms to a human". Nyrup and Robinson (2022) focus on the context of explainability: "In the conversational context C, a given phenomenon (model, system, prediction, ...), P, is explainable by an explainer, S, to an audience, A, to the extent S is able to convey information to A that enables A to draw inferences about P that are needed to achieve the purposes that are salient in C" (Nyrup & Robinson, 2022). They emphasise explanatory *pragmatism*, arguing that explanations can never be generalised, and instead must be culturally and socially situated and context-sensitive. For music, XAI should thus enable understandings of AI that are meaningful within the musical culture and its broader context, balancing human control and AI autonomy to augment – rather than replace – musicians and their tools (Shneiderman, 2022). In this chapter we focus on how to make generative AI models more interpretable so that humans, and especially musicians, can better understand and control the music generation of an AI.

There are very few XAI systems for music, and the arts in general, which means that there is very little indication of what makes a good XAI for music or how XAI might support more co-creative generative systems. Moreover, there are three key challenges to applying XAI to generative music:

(1) As outlined above, most current XAI research focuses on functional explanations of AI models (e.g. why an AI is 90% sure about its prediction), so we cannot simply apply XAI design principles to AI Music systems.
(2) The majority of XAI research offer design guidelines for explainability (Liao, Gruen, & Miller, 2020) or theories of explanation (Andres et al., 2020), but very few describe the implementation of explainable systems meaning that there are few XAI systems to build from.
(3) Possibly most importantly, using XAI in creative endeavours such as music presents a philosophical conundrum in which we must ask ourselves what it might actually mean to understand a creative AI system. For example, when musicians co-create music together they rely on intuition and non-verbal cues to manage the interaction (Healey, Leach, & Bryan-Kinns, 2005) rather than objective explanations of the rationale for each musical contribution made as may be suggested by some XAI frameworks. Given the reliance on subjective and ambiguous information in human music making we need to question what the nature of explanation in creative settings might be.

We envision that XAI could support the role of AI and generative systems as having a collaborative and iterative role in the creative process. Much like a group of musicians interacting together, there should be an element of dialogue where musicians can easily identify the AI's contributions and possibly decisions, at least from a high level. AI might function as other controllable or playable "instruments" in digital musical systems; like a musical effects plug-in in a digital audio workstation (DAW), the user might not need to understand everything down to the algorithmic level but should be able to relate what the system is contributing to their own input.

We also believe XAI can be beneficial by not dictating the creation process or a "correct" approach. AI could function in defined roles within its existing musical culture so that users know what to expect and can negotiate AI's input into existing understanding; for instance, developing and iterating on counterpoint or accompaniment parts with the user's defined melody, as an arranger might do, or offering manipulated phrases to assist in improvisation, like another member of the band. Musicians are still given agency over the creative process, without all elements being fully automated cf. HCAI (Shneiderman, 2022).

2.1. Explainable AI and Generative Music

Artificial Intelligence has been used across the culture cycle stages (Pessoa & Deloumeaux, 2009) of music from creation such as the Illiac Suite (Hiller & Isaacson, 1957) compositions, and composition experiments by the band Dadabots (Zukowski & Carr, 2018), to consumption and participation such as George Lewis's Voyager (Steinbeck, 2022) improvisations, co-creative musical agents such as Spire Muse (Thelle & Pasquier, 2021), and co-performance system such as *Shimon* the marimba playing robot (Hoffman & Weinberg, 2010) shown in figure 1. However, to date there has been no systematic review of the explainability of the AI models used in these systems. To address this we undertook a survey of 100 recent AI music papers. Publication venues reviewed included the New Instruments for Musical Expression Conference (NIME) series and the Computer Music Journal. We started with the 94 papers reviewed in Herremans, Chuan, and Chew's (2017) review and removed 19 papers which were not related to music or were inaccessible. We then added 12 papers published more recently than the Herremans et al.'s (2017) review to capture state-of-the-art (Andres et al., 2020; Benetatos, VanderStel, & Duan, 2020; Gillick & Bamman, 2021; Hazzard et al., 2019; Long, Padiyath, Teachey, & Magerko, 2021b; Louie, Coenen, Huang, Terry, & Cai, 2020a; Lupker, 2021; Malsattar, Kihara, & Giaccardi, 2019; Proctor & Martin, 2020; Roberts et al., 2019; B. L. Sturm, Santos, & Korshunova, 2015; Zhang, Xia, Levy, & Dixon, 2021), making a total of 87 papers outlined in (Bryan-Kinns et al., 2021). To complete the review for this chapter, we added an additional 13 papers (Agostinelli et al., 2023; Bougueng Tchemeube, Ens, & Pasquier, 2022a; Caillon & Esling, 2021; Dinculescu et al., 2019; Ford & Bryan-Kinns, 2022a; Huang et al., 2023; Manco, Benetos, Quinton, & Fazekas, 2021; OpenAI, 2022; Rau, Heyen, Wagner, & Sedlmair, 2022; Sarmento et al., 2023; Schneider, Jin, & Schölkopf, 2023; Vigliensoni, McCallum, & Fiebrink, 2020; Wang, 2023) to represent state-of-the-art advances since (Bryan-Kinns et al., 2021). There have been several systematic review of XAI systems e.g. Guidotti et al. (2018), but these do not consider explainability within the creative domain. Therefore we developed a classification of XAI for the arts which captures i) the role of the AI; ii) the possible interaction with the AI; iii) what stage of grounding might be inferred from the AI. These three categories draw on three existing frameworks as detailed below.

The role of the AI. We used Lubart's classification of the AI roles in creative settings (Lubart, 2005) to classify the role of the AI as follows:

- Assistant: AI which takes care of music generation without interacting with humans.
- Pen-pal: AI which engages in a back-and-forth interaction with humans e.g. taking music as an input and responding with a musical output.
- Coach: AI which provides some guidance or suggestions as part of a musical making process.
- Colleague: AI which engages in co-creation of music in a more fluid and nuanced

Figure 1. Shimon the robotic marimba player. Image courtesy of Gil Weinberg.

manner than a pen-pal.

The role AI takes must negotiate the human desire to retain freedom and influence in the creative process (Frid, Gomes, & Jin, 2020; B. L. Sturm, Ben-Tal, Monaghan, et al., 2019). Indeed, Shneiderman (2022) suggests that people feel a great pride in having control and being able to develop expertise, and is thus a crucial consideration for designing AI tools for artistic contexts such as music. Noting the dialoguic and iterative aspects of human-human collaboration within the music making process (Suh, Youngblom, Terry, & Cai, 2021), Lubart's creative AI roles mirror their human equivalents. While retaining human agency, AI can provide material and guidance within the creative process in ways that parallel "traditional" music creation. For instance, music creation in more aural traditions often involves sharing of melodic phrases and copying or modifying them for new compositions; for instance, re-using licks and recreating players' styles in jazz improvisation. For example, Irish folk songs generated by the deep learning model *folk-rnn* (B. L. Sturm & Ben-Tal, 2017) are performed and evaluated in traditional musical practice settings as illustrated in figure 2. If AI is responsible for replacing or augmenting the human's role in an aspect of dialogue, it is necessary to then evaluate it within its intended context, to see how well it fulfils that role with respect to its musical culture (B. L. Sturm & Ben-Tal, 2017; B. L. Sturm, Ben-Tal, Monaghan, et al., 2019).

Interaction with the AI. We draw on Cornock and Edmonds (1973)'s classification of interaction with interactive art to identify 3 forms of interaction with generative AI systems:

- Static: the output of the generative system does not change.
- Dynamic-Passive: the output of the generative model changes based on some

Figure 2. Performance of "The Days are Getting Longer" by folk-rnn v2 + Sturm. Image courtesy of The Society for the Preservation and Promotion of Machine Folk Music (v1.1).

data or time passing, but not in response to user input.
- Dynamic-Interactive: the output of the system changes based on user input.

The common ground with the AI. Drawing on Clark and Brennan (1992)'s categorisation of grounding in human communication we identify stages of what a person might be able infer about an AI's output state:

- Stage 0: there is no perceptible indication that the generative model has produced some output.
- Stage 1: there is some perceptible indication that the generative model has produced output.
- Stage 2: there is indication of what kind of output has been produced by the generative model.
- Stage 3: there is some indication of why the generative model produced the output that it did.

In this view the explainability of a generative AI is a combination of the role that the AI takes, the interaction that it offers to the user, and amount of grounding that a person might be able to infer about the AI's output. These three aspects are entangled views on the nature of explainability. For example, a colleague AI would involve more interaction than a pen-pal AI, and likely higher states of grounding with the AI. Similarly, establishing higher states of grounding would likely require more fluid interaction with the AI. The key point is that each of these categorisations emphasises more real-time and fluid interaction with the AI, reflecting the way in which we naturally interact with each other and the world around us.

Figure 3 summarises the results of our survey. We found that 77 of the 100 papers generated music without any human collaboration - taking an Assistant role (figure 3a). We also found that there were very few systems that offered real-time user interaction (27 of 100), with most systems generating melodies from training data with no user

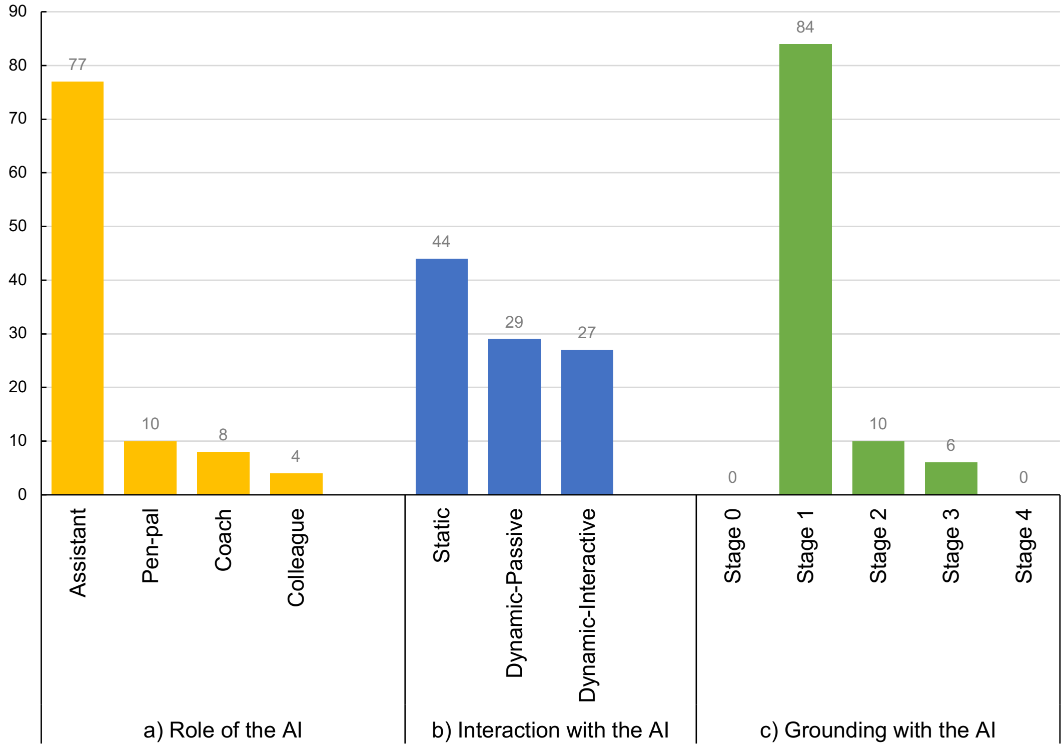


Figure 3. Survey of 100 AI Music Papers

input once the AI model was started (figure 3b). Finally, we found that most models reported in the research papers (84 of 100) offered only Stage 1 of grounding meaning that a user would only be aware that some output was generated, but not what kind of output nor why (figure 3c).

Whilst the majority of the papers we reviewed offered low levels of explainability there are a number of notable outlier models which offered more collaborative, interactive, and explainable models. For example, Shimon the robotic marimba player (Hoffman & Weinberg, 2010) provides a real-time feedback loop within the generative model as it listens to musicians in real-time and plays along in a collaborative manner, more of a colleague than a pen-pal. Furthermore, the physical movement of Shimon is designed to provide cues to the musicians on Shimon's current and possible next states, offering grounding of stage 2, though it does not explain *why* it made generative decisions (stage 3). Generative composition tools such as Hyperscore (Farbood, Kaufman, & Jennings, 2007) also offer higher levels of grounding than most papers reviewed. In Hyperscore generated music is suggested in a piano-roll notation within the same interface that musicians create their compositions. In this way user may observe the connection between their musical input and the generative AI's suggestions (stage 2 of grounding). In addition, the real-time interaction with the composition interface allows for trial-and-error modification and iteration of musical input and AI responses allowing for some intuitive understanding of the AI models to be developed (towards stage 3 of grounding) even though there is no explicit explanation of why the AI generated the music that it did.

Besides being used for music generation in transfer learning paradigms (Banar & Colton, 2021, 2022b; Payne, 2019), we found that Large Language Models such as

ChatGPT (OpenAI, 2022) work in a pen-pal role in which the AI which responds to text input by a user in a turn taking manner, and also offer some elements of coaching role when asked for advice about music composition. However, whilst such an LLM can be asked to explain aspects of its musical generation e.g. why it chose certain chord combinations, the explanations are essentially post-hoc rationalisation of the music that was generated, rather than explanations of how the generative model itself worked. Such post-hoc rationalisations appear to focus on broad stylistic aspects of the musical genres requested in the generation prompt and lacking nuance and detail which is often understood – sometimes implicitly – amongst musicians in real-world music making contexts. In this way the grounding that is offered by LLMs such as ChatGPT is at stage 2 rather than stage 3 as might intuitively be expected. Image-to-music generation systems (Banar & Colton, 2022a, 2023) where the AI responds to visual inputs to generate music function in a similar pen-pal role but are less able to offer explanation of their generative action given their use of visual stimuli for interaction.

Overall, in our survey of AI music research we found that most systems offered little explainability to users. Whilst there is perhaps a recent increase in the amount of (real-time) interaction offered to users the complexity and opaqueness of contemporary deep learning models still makes understanding what the generative AI model is doing, and why, a challenge for AI research and Human-Computer Interaction.

3. Case Study: Explaining Latent Spaces for Generative AI Music

Given the lack of explainability in generative AI music models we have been exploring the design, build, and use of a real-time generative music model based on the popular Variational Auto Encoder architecture (VAE). A VAE consists of an *encoder* which encodes training data into a multi-dimensional *latent space* which is then used by a *decoder* to generate data (music in our case) in the style of the training data as illustrated in figure 5. A key aspect of VAEs for interactive applications is that modifying values of the latent space dimensions will have an effect on the generated data which are in keeping with the training data. For musical applications this means that users can navigate around a latent space to generate variations of music in the style of the training data e.g. (Louie, Cohen, Huang, Terry, & Cai, 2020; Murray-Browne & Tigas, 2021b; Pati, Lerch, & Hadjeres, 2019; Thelle & Pasquier, 2021; Vigliensoni et al., 2020). For example, *Sonified Body* (Murray-Browne & Tigas, 2021a, 2021b) transforms a dancer's movement into sound using the latent space of a VAE trained on individual dancer's movements. However a key challenge for explainable VAE systems is that unsupervised learning is used to create the latent space which means that whilst users can navigate around the space there may be no meaningful connection between the movement in the latent space and the music produced resulting in low levels of grounding. Moreover, retraining such models leads to different mappings (Murray-Browne & Tigas, 2021a) making such latent spaces additionally opaque. Supervised regularisation methods such as latent space regularisation (LSR) (Hadjeres, Nielsen, & Pachet, 2017) have been used in training VAEs for user controlled generation of images (Lample et al., 2018) and music (Pati & Lerch, 2019; Tan & Herremans, 2020) offering more meaningful control of VAE based music generation through increased structure and labelling of the models. In this chapter we explore how both interaction and grounding can can be increased within one real-time generative system. For us, the combination of real-time interaction coupled with meaningful exposure of the AI

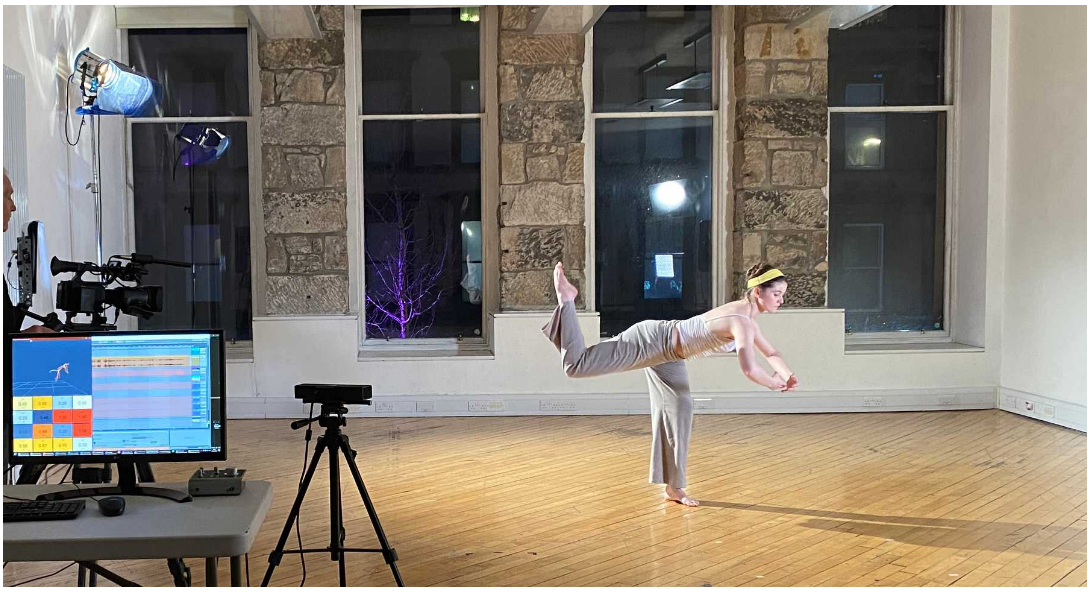


Figure 4. Sonified Body. Image courtesy of Tom Murray-Browne.

model is key to increased explainability.

We built our generative AI system (Bryan-Kinns et al., 2021) on the popular MeasureVAE system (Pati & Lerch, 2019)[1] in which a measure of music is represented as 24 characters including notes and rests. We trained it using publicly available music from a single genre in keeping with current AI music generation research - in our case 20,000 publicly available monophonic Irish folk melodies (B. L. Sturm, Santos, Ben-Tal, & Korshunova, 2016). This training produces a latent vector of 256 latent dimensions. To increase the explainability of the latent space we follow Pati and Lerch (2019, 2020) and use LSR to force some dimensions of the latent space to represent musical attributes. Specifically, we use meaningful musical features used in current research following (Pati & Lerch, 2019, 2020) to demonstrate our approach:

- Rhythmic Complexity (RC): Toussaint (2002)
- Note Range (NR): *maximum pitch minus minimum pitch*
- Note Density (ND): *number of notes in a measure*
- Average Interval Jump (AIJ): *the average of the absolute values of the interval between adjacent notes in a single measure.*

We assign these attributes to latent dimensions *0*, *1*, *2* and *3* respectively - see Figure 6. See (Pati et al., 2019) for a full description of MeasureVAE, and see (Bryan-Kinns et al., 2021) for full details of our implementation.

For training of our model, as detailed in (Bryan-Kinns et al., 2021), we use use Adam (Kingma & Ba, 2014) as the optimizer with learning rate = 1e-4, β_1 = 0.9, β_1 = 0.999 and ϵ = 1e-8. Training the model on a single GeForce RTX 2080 Ti GPU takes an average of 2.5 hours per epoch for a total of 30 epochs following (Pati & Lerch, 2019). The model achieves 99.87% reconstruction accuracy on the training set and 99.68% accuracy on the validation set. The interpretability scores which measure how well we can predict an attribute using only one dimension in the latent space

[1] `https://github.com/ashispati/AttributeModelling` licensed under Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

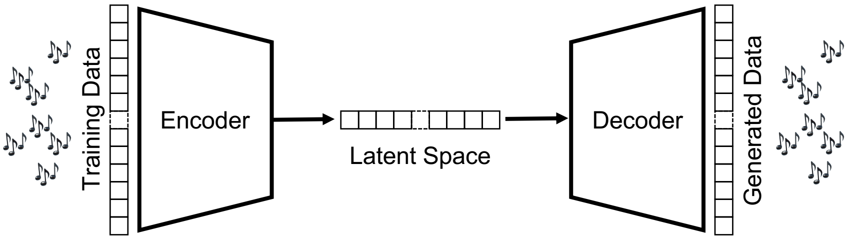


Figure 5. A simplified Variational Auto Encoder architecture

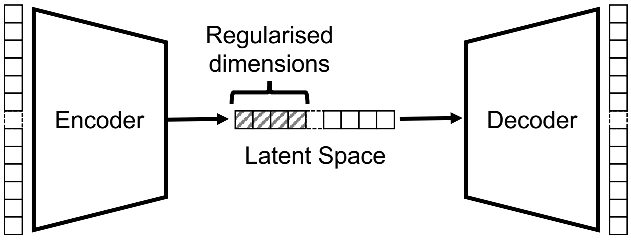


Figure 6. The simplified MeasureVAE with LSR

(Adel, Ghahramani, & Weller, 2018) average 0.92 (the closer to 1.0 the better): RC 0.80, NR 0.99, ND 0.99 and AIJ 0.91.

To support real-time interaction with our generative model we built a user interface (UI) as a web application[2] using React.js to demonstrate the interaction as illustrated in figure 7. To provide real-time generative performance we sweep the four regularised latent dimensions (representing the 4 musical attributes) and discretely take sample values to generate 10,000 latent vectors for the 4 regularised dimensions. These are then decoded into music sequences which are used to generate MIDI files, piano-rolls, and MP3 files in advance. These pre-generated files are available online[3]. By combining real-time interaction with more meaningful control using LSR we aim to increase the explainability of our generative system.

The web app shows the input MIDI measure on the lefthand side of the UI, with the generated output measure show on the righthand side (figure 7). The MIDI measures can be listened and provide a way to *audition* the music. The main area of interaction for the web app are the two 2D pads in the centre of the UI. These provide a visualisation of two pairs of regularised dimensions which also allow users to navigate within the latent space and training sets. By moving the red dot around the 2D visualisation the generated MIDI output is updated in real-time providing an interaction loop between user and generative AI. Different pairs of dimensions can be visualised in the 2D pads using the toggle buttons above the 2D pads allowing user to navigate across different combinations of regularised dimensions - the left pad navigates Rhythmic Complexity and Note Range whereas the right pad navigates Note Density and Average Interval Jump. The visualisation also indicates where the input MIDI is located in the latent space (not illustrated in the figure).

We suggest that our web app increases the explainability of the generative AI in two ways as discussed below. Firstly it exposes the latent space of the generative model to users in meaningful ways through visualisation and latent semantic regularisation

[2]`https://xai-lsr-ui.vercel.app/`

[3]`https://github.com/bbanar2/Exploring{_}XAI{_}in{_}GenMus{_}via{_}LSR`

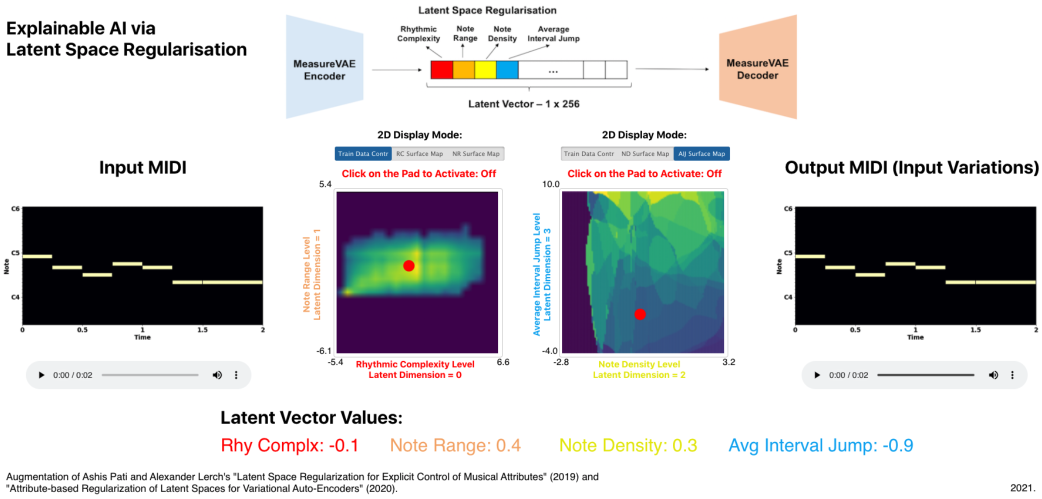
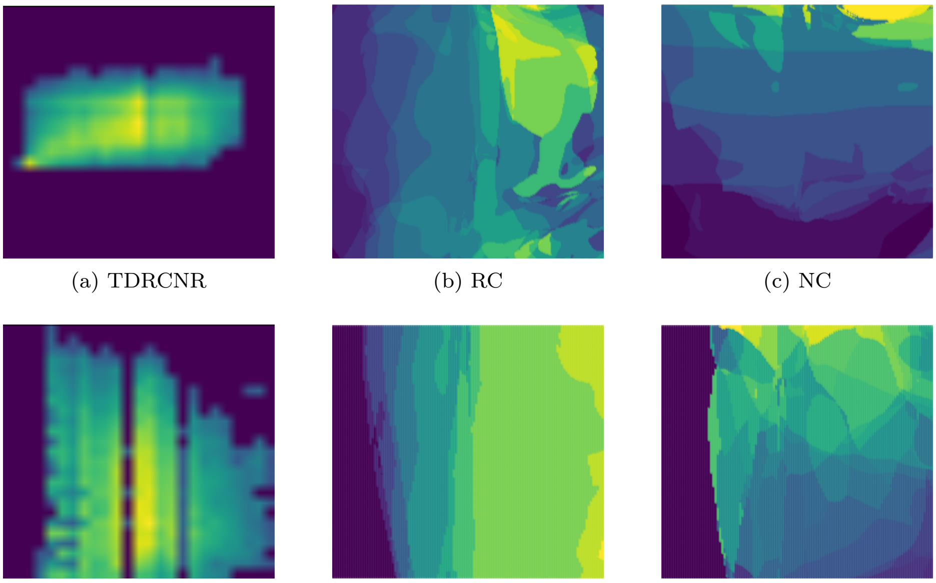


Figure 7. Screenshot of the web application user interface from Bryan-Kinns et al. (2021)

using musical features. Secondly, the real-time interaction of the UI allows for users to develop an understanding and intuition of how the AI model generates music in relation to the input music and the control parameters.

3.1. Visualisation

The web app displays two visualisations of the training data and latent space in the centre of the UI based on the visualisations by (Pati & Lerch, 2019) as illustrated in figure 8. The left hand plot in the UI can be toggled to show one of 8a, b, or c. Similarly, the user can select one of 8d, e, or f to be shown in the right hand visualisation in the UI. The training data plots (8a and 8d) indicate how many items in the training set have contributed to the selected location in the latent space with brighter colours indicating more items. The visualisations of the latent space 8b, c, e, and f show the musical attribute values for our regularised dimensions e.g. 8b shows the Rhythmic Complexity (RC) of music generated by plotted across all values of latent dimension 0 (RC) and 1 (NR) as a 2D plot. Again, brighter colours indicate higher values. These surface maps give an indication of how the regularisation technique works and suggest to users areas in which to explore different musical features. As movement in the visualisations generates music in real time the surface maps offer grounding stage 2 - provide an indication of the kind of music that was produced in response to user input. Furthermore, the visualisation offers an opportunity for users to learn about the mappings from latent space to music generation and to start to develop inference about possible kinds of music produced in different areas of the maps. This then moves the explainability of the AI from stage 2 towards stage 3 where the might develop some understanding of why the AI generates the music it does for different parts of the latent space. An explanation of the link between latent space and music generated is not made explicitly, but rather there is the potential for users to infer an understanding of this link through interaction with the visualisation. By providing visualisations of the latent space we move beyond the explainability of other generative models which allow users to interact with latent space but do not visualise the space itself e.g. (Murray-Browne & Tigas, 2021b; Thelle & Pasquier, 2021; Vigliensoni et al.,



(a) TDRCNR (b) RC (c) NC

(d) TDNDAIJ (e) ND (f) AIJ

Figure 8. Visualisations from (Bryan-Kinns et al., 2021): a) Training Data visualised in terms of rhythmic complexity and note range (TDRCNR); b) Rhythmic Complexity surface map (RC); c) Note Range surface map (NR); d) Training Data visualised in terms of note density and average interval jump (TDNDAIJ); e) Note Density surface map (ND); f) Average Interval Jump surface map (AIJ)

2020).

3.2. Real-time Interaction

The real-time nature of the interaction with our UI, particularly the real-time generation of music in response to navigation within the latent space may offer users ways to develop an intuition about the working of the model. In other words, the immediate feedback and interaction loop offer opportunities for users to be able to get the *gist* of how the AI works in terms of the relationship between the latent space and the music it may produce when the musical attributes are changed. Moreover, the UI keeps the original input used to generate music and shows the original location in the latent space whilst the latent space is being navigated. We suggest that this allows for exploratory or trial-and-error learning of the AI model as the user can always return to the original state. It also allows for contrastive example based learning as the user can compare their original input to the variety of generative outputs produced as the space is navigated. There are similarities here to the design of Hyperscore (Farbood et al., 2007) where the original melody is always shown as reference whilst whilst musical parameters are changed. Our real-time interaction approach also aligns with creativity support tool design principles such as (Shneiderman et al., 2006)'s principles of supporting explorations and designing creativity support tools which offer a high ceiling for creativity.

The real-time feedback loop in combination with the visualisation of latent space and a user's location within it offers chances for increasing grounding with the AI from a stage 1 where we know that the AI has done something, but we are not sure what, to stage 2 where users may start to build an understanding of what the AI has done in response to their input.

In terms of the role of our AI, the real-time interaction moves our model from being a pen-pal to more of a colleague interacting in real time, though there is a lack of agency on the part of our model to really co-create with. The model also lacks any explicit explanation of why it generates the music it does - the user must infer the explanation from their navigation of the latent space, meaning that the model does not take on the role of coach. In terms of music making it is plausible that this makes the model more suited to a duet style improvisation for real-time performance, or a composition idea generator for less live situations.

4. XAI for Music Challenges

Building our generative AI model and reflecting on its use has raised a number of research challenges for explainable AI and music outlined in this section. The challenges have been used to inform thinking about explainable AI and the arts more broadly (Bryan-Kinns, Ford, et al., 2023).

4.1. Challenge: The Nature of Explanation

The open ended nature of creative endeavours such as music making raises questions about how much we might want to understand and AI in creative settings and what kinds of explanations might be appropriate. Being surprised, confused, delighted, or reflective (Candy, 2019; Ford & Bryan-Kinns, 2022b, 2023) are often integral to creative activities. Indeed, failure is often embraced as a resource in creative endeavours (Hazzard et al., 2019). These characteristics of creative activities require a balance between skills and challenge (cf. flow theory Csíkszentmihályi (1990)), for example, balancing between explainability and surprise, efficiency and serendipity, ease-of-use and playful challenge. These balances are in contrast to the more functional goals of explainability, transparency, and interpretability outlined in definitions of XAI proposed in (Gunning, 2016) and expanded in (Guidotti et al., 2018), and raise questions about what the nature of explainability in creative settings might be. For example, in human group music improvisation most of the communication that hold the group together is non-verbal and implicit, such as gestures, nods, physical position, eye contact, or even musical emphasis (Healey et al., 2005). The musicians in such settings do not request an in-depth explanation of each note that is produced and why it was chosen but instead rely on an intuition about each other's performance styles and practice to understand how the improvisation might progress within the socially constructed setting of music making. Generative AI systems such as Shimon (Hoffman & Weinberg, 2010) and McCormack et al. (2019)'s collaborative AI drummer embrace this balance between communication of system state and the socially constructed expectations of music improvisation by communicating indications of current potential future AI states through physical gesture and emoticons respectively. In this way the explanations are much more about conveying the *gist* of what the AI is doing and why, and very much less about providing in-depth explanations of features of models or predictions. Partly this is due to the in-the-moment nature of music making which requires sufficient information to progress the collaboration, but not so much that it stifles the creative process itself.

4.2. Challenge: AI Models, Features, and Training Data

There are an ever increasing number of generative AI models for music and the arts more broadly. For music generation the models range from probability based models such as Markov Chains (Ames, 1989; Whorley & Laney, 2020) through to deep learning techniques (Briot & Pachet, 2018). Whilst probablistic approaches typically offer more explainable models than deep learning models their outputs are often less novel than deep learning models. In this chapter we chose to use the MeasureVAE architecture to build our demo as such deep learning models have become increasingly popular in recent years and generate convincing musical outputs (Carnovalini & Rodà, 2020; Herremans et al., 2017; B. L. Sturm, Ben-Tal, Úna Monaghan, et al., 2019). We could have chosen other models and architectures for our demo and this would have changed the nature of the explanations that could be imposed on and offered by the model. However there is very little research about how AI models are used in creative practice (Karimi, Rezwana, Siddiqui, Maher, & Dehbozorgi, 2020; Louie, Cohen, et al., 2020; B. Sturm, 2022), and we found no research about how XAI is used in music making. This means that there are open research questions about which XAI models are more or less appropriate for different forms of artistic practice broadly such as music or visual arts, and which are appropriate for different activities within an art form such as music composition and music improvisation.

In addition to a lack of research on which XAI models might be appropriate for music making we found no research on the suitability or appropriateness of features used in explanations. For example, in this chapter we selected four musical parameters to be used in our XAI system - rhythmic complexity, note density, note range, and average interval jump. These were selected as representative of current generative AI research and Music Information Retrieval research, but they are only a small subset of the possible musical features we could have used, and we do not know what effect the different features chosen have on the performance of the model or its explanations. Moreover, the features we chose focused on the pitch and rhythmic aspects of a measure of music and did not consider higher level features such as timbre, genre, mood, composition structure, and so on which musicians may be interested in when making music. In addition, for a non-musical user, the system may still present some explainability barriers as the chosen parameters – although labelled and visualised – require the user to have some prior understanding of music and musical features. Other control mechanics, such as the semantic sliders used in (Louie, Cohen, et al., 2020), may be more appropriate here, although these currently do not give explanations for the link between interaction and output.

In keeping with other generative music research we trained our demo system on a commonly used Irish Folk Music dataset (B. L. Sturm et al., 2016). However, this limits the tonal and rhythmic features of the music we generate to the characteristics of the Western music canon, specifically folk music. The reliance on datasets such as the Irish Folk Music dataset mean that generative systems such as ours and others in the field reinforce the marginalisation of smaller datasets and other genres leading to limited diversity of musical styles used and generated. Future research needs to consider how musical genres beyond the Western canon can be better included in AI training, especially where model training might require large datasets biasing them towards using large and well established datasets which are typically within the Western canon. More broadly, this challenge of diversifying training data connects with one of the grand challenges of Human-Centred AI (Garibay et al., 2023) which is to design AI that is inclusive by using de-biased datasets to ensure fairness and to promote accountability by developers.

To begin to explore these challenges of XAI models, features, and training datasets we compared the MeasureVAE model with 4 regularised dimensions used in this chapter to an AdversarialVAE (Kawai, Esling, & Harada, 2020) with the same 4 musical features applied to control its music generation. We compared the effect of different numbers of latent dimensions on music generation, and compared the effect of training MeasureVAE on datasets of four different musical genres: Irish folk, Turkish folk, Classical, and pop. In this comparison (Bryan-Kinns, Zhang, Zhao, & Banar, 2023) we found that MeasureVAE was better at generating music in the style of its training data whereas AdversarialVAE was more controllable with the four musical attributes. We also found that MeasureVAE performed best when generating low complexity music such a pop and rock, and that 32 or 64 latent dimensions were optimal for MeasureVAE when regularising four of the dimensions to the musical features used in this chapter. This initial investigation illustrates the impact that AI models, datasets, and musical features have on generative models which are designed to be explainable and highlights the importance of exploring these aspects in future XAI research.

4.3. Challenge: User Centred Design for XAI

The challenges above highlight the need for involving users in the design of XAI for music. For example, to ensure that appropriate attributes for explanation are selected for musicians to actually use in their practice, and more broadly to explore how artistic identity can be incorporated and expressed in, with, and through such systems. Involving users in AI design and co-design a grand challenge for HCAI research (Garibay et al., 2023) and yet there is very little research on using User Centred Design (UCD) approaches to drive the creation of XAI systems (Zhu, Liapis, Risi, Bidarra, & Youngblood, 2018). Moreover, there is no research to date on UCD for XAIxArts. This means that there are open research questions about how to include musicians in the co-design of generative music systems. Similarly, there are open research questions about how such XAI systems are used, unused, misused, and appropriated in real-world music making practice. Indeed, whilst there have been reports of the use of AI in music making such as the Illiac Suite (Hiller & Isaacson, 1957), Calliope (Bougueng Tchemeube, Ens, & Pasquier, 2022b), or experiments by the band Dadabots (Zukowski & Carr, 2018)) there is no research to date on the role of explainable AI in such creative settings. Furthermore, as highlighted earlier, there is a need for XAI designers to develop systems which exhibit sensitivity towards the particular social and cultural context (Nyrup & Robinson, 2022), especially the case in music which often relies on hyper-localised references and traditions.

It should be noted that in this chapter we have focused primarily considered individual musicians as users of generative XAI systems. Other possible XAI target audiences are also important to consider such as ensembles of musicians, and audiences of music performances, who could equally stand to benefit from clearer explanations of generative AI in music performance. Taking a broader and more critical and post-human perspective, XAIxArts interfaces could follow calls in HCI research to subvert user-centricity in favour of more-than-human approaches (Wakkary, 2021).

To begin to explore how such XAI systems might be appropriated in existing music making practice we developed a second UI (Banar, Bryan-Kinns, & Colton, 2023) for the demo outlined in this paper. This second UI was built as a Max4Live plugin to allow deployment and use within musicians' musical practice and music making

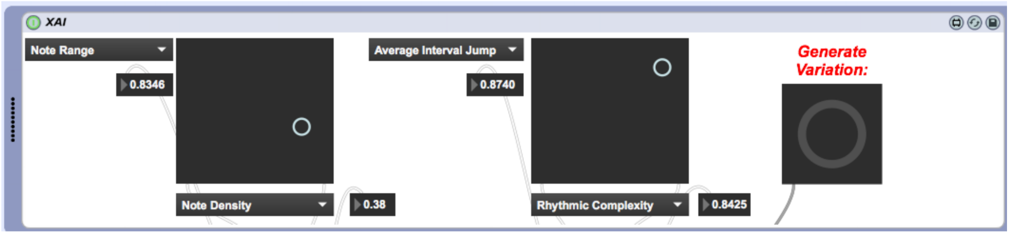


Figure 9. Screenshot of the Max4Live plugin user interface

workflows as illustrated in figure 9. To explore its use in music making practice an autoethographic study incorporating the demo into a rule-based composition practice and music workflow was carried out (Noel-Hirst & Bryan-Kinns, 2023). In this way the demo was deployed beyond its original genre (Irish Folk) and expanded its use beyond a standalone music generator by integrating it into a music workflow including a rule-based rhythm generator, monophonic processor, piano roll, and MeasureVAE demo. In the study the combination of rule-based composition systems with generative AI was found to offer new ways to musically *explore* the latent space of the generative model - by permuting the output of the rule-based system and using this to drive navigation of the latent space the musician-researcher was able to "explore the under-defined contours of the latent space through example" (ibid.). In this way music generated by MeasureVAE was driven and explored by a rule-based composition system rather than by direct manipulation of parameters. The study also highlighted the potential of the XAI to provide *insight* into more opaque systems. In this case the musician-researcher used the visual representation of the latent space in the UI as a way of understanding the output of their rule-based composition system which in turn offered opportunities for fine grained control of the rule-based system not easily achieved without the visualisation. Whilst these are preliminary findings from early stage research they already highlight the potential use and appropriation of such a XAI music system beyond being a standalone Irish folk music generator.

4.4. Challenge: Interaction Design of XAI

There is little research to date on how to design the interaction with XAI for music and the arts more broadly. For our demo this raised Interaction Design questions including: which AI features to present to the user; how to visualise and navigate high dimensional data, in our case 4 regularised dimensions, but there could very well have been many more dimensions to interact with; how to represent and manage the entanglement between dimensions; and how to provide real-time interaction in a temporal art form (music). These Interaction Design challenges will be compounded when designing for use in music practice and real-world workflows where interacting with the AI will likely be a secondary focus to the overall creative practice and workflow.

Visualisation is currently the primary design discipline of interest with regards to XAI, and there are many surveys of visual design methods for XAI that can serve to ground the development of XAI for music and XAIxArts interfaces more broadly. Alicioglu and Sun (2022) provide a comprehensive survey of such methods, noting several important challenges. The main challenge is scalability in data representation as dimensionality increases, both from a visual clutter perspective and from an interface performance perspective. This has always been a key issue for musical interface design, which is in this case compounded by the curse of dimensionality. A close second is that despite the extent of XAI research, there is no consensus and little evaluative literature on which data and models should be paired with which visual methods; instead they find that most solutions trend towards specialisation. This point is reiterated in another recent review by Krajna, Kovac, Brcic, and Šarčević (2022), who note that there is similarly a lack of standardised dataset and unbiased metrics for evaluating XAI systems and interfaces. They also note an additional challenge, that XAI research is constantly playing catch-up to the latest developments in model architectures and data regimes, with no guarantees that existing XAI methods will necessarily transfer to new innovations.

One area of interactive visualisation research that can provide designers with useful examples and inspirations is in *interactive articles*, which are sometimes known as *explorable explanations*[4] after the Bret Victor article of the same name[5]. The *explorables* community of data visualisation designers, data journalists and open source software developers has explored many approaches to using narrative, plurality of representations, and rich multimedia to enhance comprehension of complex phenomena of many kinds (Heer, Kale, & Jeffrey, 2019). This community directly influenced Distill[6], an experimental publication platform that ran from 2016-2021, dedicated to innovation in making machine learning research "clear, dynamic and vivid".

Aside from the visual methods, auditory display (Kramer, 1993), auditory augmentation (Weger, Hermann, & Höldrich, 2018, 2022), sonification (Hermann, Hunt, & Neuhoff, 2011) and sonic interaction design (Franinovic & Serafin, 2013) are unexplored with regards to XAI for music, and are clear choices for taking advantage of the listening skills of musicians. These domains are already being explored for increasing understanding of AI, for example in *AIive* which introduces interactive visualisation and sonification of neural networks in virtual reality (Lyu, Li, & Wang, 2021), and *SonOpt* which proposes methods for sonification of bi-objective population-based optimisation algorithms (Asonitis, Allmendinger, Benatan, & Climent, 2022). Toolkits for sonification are also becoming more widely accessible and integrated into machine learning contexts, for example *sc3nb*, a new Python-based sonification framework (Hermann & Reinsch, 2021). The sonification literature also spans beyond informative displays and technical innovations, considering the impact of aesthetic and ritualisatic dimensions of sonic communication Morabito, Armitage, and Magnusson (2022), all of which can be drawn on to inspire.

The challenges and opportunities in interaction design for XAI for music are not constrained only to the visual and audio domains. Music is in many ways also a tactile domain, and Maliheh Ghajargar *et al.* have in recent years pioneered what they refer to as *graspable AI* (Ghajargar & Bardzell, n.d.; Ghajargar, Bardzell, Smith-Renner, Höök, & Krogh, n.d.). Their approach centres around the exploration of different forms of embodiment and tangibility for making AI understandable, building off the rich HCI literature in this domain to meet contemporary challenges around XAI, FAccT and HCAI. Music interaction designers and researchers are of course well aware of the potential affordances of haptics (Papetti & Saitis, 2018). Tangible XAI is particularly promising in that it hints at approaches that avoid some of the aforementioned challenges with visual methods, while also tapping into extant musical sensorimotor skill. It also proposes to build off less explicit, subconscious intuition acquisition processes associated with direct manipulation (Nimkulrat, 2020), that are also being exploited

[4] `https://explorabl.es`
[5] `http://worrydream.com/ExplorableExplanations/`
[6] `https://distill.pub/`

in craft HCI contexts to enable manual exploration of subtle, micro scale details of digital artifacts (Armitage, Magnusson, & McPherson, 2023a, 2023b).

We also wish to highlight conceptual approaches to interaction design for XAI for the arts that may prove provocative and perhaps useful. Sanchez *et al.* (Sanchez, Caramiaux, Thiel, & Mackay, 2022) found that exposing model uncertainty to users in an interactive machine learning (IML) context increased user confidence. While datasets and models have greatly increased in size and complexity of late, researchers still turn to toy models (Chughtai, Chan, & Nanda, 2023) and synthetic data (Liu, Michaud, & Tegmark, 2023) to reveal fundamental insights. These approaches show that getting hands-on with the exact opposite of the complex black boxes users are commonly presented with today, can be a driver for intuition and trust. Interaction designers should explore not just synthesised explanations, but also training processes, synthetic datasets and simplified but more interactive versions of entire models.

The XAI demo discussed in this chapter follows HCAI principles in giving the user more control over how they navigate a data set, whilst the AI automates the writing of measures of music (Shneiderman, 2022). A potential challenge, to be explored in future user studies, is whether users blindly trust that the systems output will fit into their own creative practice. HCAI guidelines on responsible design (Garibay et al., 2023) suggest that AI designs should actively dissuade users from blind trust of a machine and instead persuade them to critically question AI output. However, in music we have seen tools perceived as infallible and their use as being the “best” option because of cultural perceptions of technology somehow knowing better than human understanding (Reed & McPherson, 2023). Much of technology use in day-to-day life revolves around providing objective output; culturally, we are predisposed to believing tech interventions provide one correct response. Perhaps, approaches to designing for reflection (Ford & Bryan-Kinns, 2023) will be useful to explore, moving away from approaches on designing engagement where self-awareness and conscious thinking is discouraged cf. flow theory (Csíkszentmihályi, 1990), which is more frequent in classic creativity support tool research (Shneiderman et al., 2006). Indeed, it may be that XAI should be aiming to ensure that people trust themselves creatively as much as the tool itself. Endeavors in explainable AI might then focus on shifting AI’s perceived role as providing “optimal” or “correct” solutions to that of just another tool in the creative arsenal.

5. Conclusions

Explainable AI and Human-Centred AI more broadly are growing research fields which have the potential to contribute to making AI systems more artist-centric. In doing so XAI and HCAI for the arts would offer opportunities to augment and supercharge human creativity and agency rather than dis-empowering and disenfranchising in a world of autonomous AI generation systems. However, as highlighted in this chapter, current AI models for music generation offer very little explanation and human-centred interaction, typically generating content with little or no user-interaction. As AI systems are increasingly used for music creation they inevitably shape our musical culture, increasing the need for explainable AI for music to help us understand their use and behaviour. Indeed, increasingly autonomous AI risks diminishing human creativity, removing human agency from creative processes, and devaluing human creativity itself. As noted in this chapter, the reliance on a small set of training datasets predominantly from the Western canon also risks the homogenisation of music to this narrow set of genres and a marginalisation of musical skills and traditions which not in our current training sets and which may not be amendable to generation by current AI models.

In this chapter we highlighted the lack of explainability in current AI models for music. We proposed a classification framework for XAI for the arts which considers the role of the AI, the interaction offered by the AI, and the potential grounding with the AI in interaction. We suggest that designing for increasing levels of the entangled categorises increases the explainability of AI in creative settings. We demonstrated the development and use of an XAI for music by offering musical control of music generated by an AI. By comparing our model, features, and training dataset to others we illustrate the impact of these on the performance of XAI models.

Our view is that XAI for music and the arts more generally needs to move away from functional explanations deployed in other domains and focus more on conveying the *gist* of what the AI is doing and why. This move of XAI from functional domains to more creative, aesthetic, and experiential domains mirrors in some ways the shift of HCI from usability to third wave HCI (Bødker, 2015) concerns of meaning making and user experience. As we move into these more creative and cultural domains our XAI research also needs to embrace the creative practice, norms, and cultural dimensions that socially construct art.

Acknowledgements

We wish to thank Bingyuan Zhang, Songyan Zhao, Ashley Noel-Hirst, and Simon Colton who previously contributed to work (Bryan-Kinns et al., 2021; Bryan-Kinns, Zhang, et al., 2023) discussed in this chapter.

Funding

This research was funded in part by the Queen Mary University of London Research Enabling Fund. Corey Ford is a research student at the UKRI Centre for Doctoral Training in Artificial Intelligence and Music, supported by UK Research and Innovation [grant number EP/S022694/1].

References

Adel, T., Ghahramani, Z., & Weller, A. (2018, 10–15 Jul). Discovering Interpretable Representations for Both Deep Generative and Discriminative Models. In J. Dy & A. Krause (Eds.), *Proceedings of the 35th International Conference on Machine Learning* (Vol. 80, pp. 50–59). PMLR. Retrieved from `https://proceedings.mlr.press/v80/adel18a.html`

Agostinelli, A., Denk, T. I., Borsos, Z., Engel, J., Verzetti, M., Caillon, A., … Frank, C. (2023). *Musiclm: Generating music from text.*

Alicioglu, G., & Sun, B. (2022). A survey of visual analytics for Explainable Artificial Intelligence methods. *Computers & Graphics*, *102*, 502–520.

Ames, C. (1989). The markov process as a compositional model: A survey and tutorial. *Leonardo*, *22*, 175 - 187.

Andres, J., Wolf, C. T., Cabrero Barros, S., Oduor, E., Nair, R., Kjærum, A., … Madsen, B. S. (2020). Scenario-based XAI for Humanitarian Aid Forecasting. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–8).

Armitage, J., Magnusson, T., & McPherson, A. (2023a). A Scale-Based Ontology of Digital Musical Instrument Design. In *Proc. New Interfaces for Musical Expression.* Mexico City, Mexico.

Armitage, J., Magnusson, T., & McPherson, A. (2023b). Studying Subtle and Detailed Digital Lutherie: Motivational Contexts and Technical Needs. In *Proc. New Interfaces for Musical Expression.* Mexico City, Mexico.

Asonitis, T., Allmendinger, R., Benatan, M., & Climent, R. (2022). SonOpt: Sonifying Bi-objective Population-Based Optimization Algorithms. In T. Martins, N. Rodríguez-Fernández, & S. M. Rebelo (Eds.), *Artificial Intelligence in Music, Sound, Art and Design* (pp. 3–18). Cham: Springer International Publishing.

Banar, B., Bryan-Kinns, N., & Colton, S. (2023). A Tool for Generating Controllable Variations of Musical Themes Using Variational Autoencoders with Latent Space Regularisation. In *Proceedings of 37th AAAI Conference on Artificial Intelligence.*

Banar, B., & Colton, S. (2021). Generating Music with Extreme Passages Using GPT-2. In *EvoStar 2021.*

Banar, B., & Colton, S. (2022a). Connecting Audio and Graphic Score Using Self-supervised Representation Learning - A Case Study with Gyorgy Ligeti's Artikulation. In *International Conference on Computational Creativity (ICCC).*

Banar, B., & Colton, S. (2022b). A Systematic Evaluation of GPT-2-Based Music Generation. In *International Conference on Computational Intelligence in Music, Sound, Art and Design (EvoMUSART) (Part of EvoStar)* (pp. 19–35).

Banar, B., & Colton, S. (2023). A Tool for Composing Music via Graphic Scores in the Style of Gyorgy Ligeti's Artikulation Using Self-supervised Representation Learning. In *The AAAI-23 Workshop on Creative AI Across Modalities.* Retrieved from https://openreview.net/forum?id=c7W3iTr693

Benetatos, C., VanderStel, J., & Duan, Z. (2020). BachDuet: A Deep Learning System for Human-Machine Counterpoint Improvisation. In *Proc New Interfaces for Musical Expression (NIME'20).*

Benyon, D., & Macaulay, C. (2002, 07). Scenarios and the HCI-SE design problem. *Interacting with Computers*, *14*, 397–405. Retrieved from https://doi.org/10.1016/S0953-5438(02)00007-3

Bødker, S. (2015, aug). Third-wave HCI, 10 years later—participation and sharing. *Interactions*, *22*(5), 24–31. Retrieved from http://dl.acm.org/citation.cfm?doid=2818696.2804405

Bougueng Tchemeube, R., Ens, J. J., & Pasquier, P. (2022a). Calliope: A co-creative interface for multi-track music generation. In *Creativity and cognition* (p. 608–611). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3527927.3535200

Bougueng Tchemeube, R., Ens, J. J., & Pasquier, P. (2022b). Calliope: A co-creative interface for multi-track music generation. In *Creativity and cognition* (p. 608–611). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3527927.3535200

Briot, J.-P., & Pachet, F. (2018). Deep learning for music generation: challenges and directions. *Neural Computing and Applications*, *32*, 981-993.

Bryan-Kinns, N., Banar, B., Ford, C., Reed, C. N., Zhang, Y., Colton, S., & Armitage, J. (2021). Exploring XAI for the Arts: Explaining Latent Space in Generative Music. In *Proceedings of the 1st Workshop on eXplainable AI Approaches for Debugging and Diagnosis (XAI4Debugging@NeurIPS2021).* Retrieved from https://xai4debugging.github.io/files/papers/exploring_xai_for_the_arts_exp.pdf

Bryan-Kinns, N., Ford, C., Chamberlain, A., Benford, S., Kennedy, H., Li, Z., ... Rezwana, J. (2023). Explainable AI for the Arts: XAIxArts. In *Proceedings of ACM Creativity and Cognition conference.*

Bryan-Kinns, N., Zhang, B., Zhao, S., & Banar, B. (2023). Exploring Variational Auto-Encoder Architectures, Configurations, and Datasets for Generative Music Explainable AI.

In review: Machine Intelligence Research.
Caillon, A., & Esling, P. (2021). *Rave: A variational autoencoder for fast and high-quality neural audio synthesis.*
Candy, L. (2019). *The creative reflective practitioner.* Routledge.
Carnovalini, F., & Rodà, A. (2020). Computational Creativity and Music Generation Systems: An Introduction to the State of the Art. *Frontiers in Artificial Intelligence*, *3*, 14. Retrieved from https://www.frontiersin.org/article/10.3389/frai.2020.00014
Chughtai, B., Chan, L., & Nanda, N. (2023, February). *A Toy Model of Universality: Reverse Engineering How Networks Learn Group Operations* (No. arXiv:2302.03025). arXiv.
Ciatto, G., Schumacher, M. I., Omicini, A., & Calvaresi, D. (2020). Agent-Based Explanations in AI: Towards an Abstract Framework. In *Explainable, Transparent Autonomous Agents and Multi-Agent Systems: Second International Workshop, EXTRAAMAS 2020, Auckland, New Zealand, May 9–13, 2020, Revised Selected Papers* (p. 3–20). Berlin, Heidelberg: Springer-Verlag. Retrieved from https://doi.org/10.1007/978-3-030-51924-7_1
Clark, H., & Brennan, S. (1992). Grounding in Communication. *Archives*, *7*(July), 734–738.
Colton, S., & Wiggins, G. A. (2012). Computational Creativity: The Final Frontier? In *Proceedings of the 20th European Conference on Artificial Intelligence* (pp. 21–26). NLD: IOS Press. (event-place: Montpellier, France)
Cornock, S., & Edmonds, E. (1973). The Creative Process Where the Artist Is Amplified or Superseded by the Computer. *Leonardo*, *6*(1), 11–16. Retrieved from http://www.jstor.org/stable/1572419
Csíkszentmihályi, M. (1990). *Flow: The Psychology of Optimal Experience.* New York, USA: Harper Collins.
De Man, B., Reiss, J., & Stables, R. (2017). Ten years of automatic mixing. In *Proceedings of the 3rd workshop on intelligent music production.*
Dinculescu, M., Engel, J., & Roberts, A. (Eds.). (2019). *Midime: Personalizing a musicvae model with user data.*
Du, N., Haspiel, J., Zhang, Q., Tilbury, D., Pradhan, A. K., Yang, X. J., & Robert Jr, L. P. (2019). Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload. *Transportation research part C: emerging technologies*, *104*, 428–442.
Farbood, M., Kaufman, H., & Jennings, K. (2007). Composing with Hyperscore: an Intuitive Interface for Visualizing Musical Structure. In *Proceedings of the 2007 International Computer Music Conference, ICMC 2007, August 27-31, 2007.* Michigan Publishing.
Ford, C., & Bryan-Kinns, N. (2022a). On the role of reflection and digital tool design for creative practioners. In *Digital skills for the creative practitioner: Supporting informal learning of technologies for creativity workshop at chi 2023.*
Ford, C., & Bryan-Kinns, N. (2022b). Speculating on reflection and people's music co-creation with ai. In *Workshop on generative ai and hci at the chi conference on human factors in computing systems 2022.* Retrieved from https://qmro.qmul.ac.uk/xmlui/handle/123456789/80144
Ford, C., & Bryan-Kinns, N. (2023). Towards a Reflection in Creative Experience Questionnaire. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3544548.3581077
Franinovic, K., & Serafin, S. (2013). *Sonic interaction design.* MIT Press.
Frid, E., Gomes, C., & Jin, Z. (2020). Music Creation by Example. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (p. 1–13). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3313831.3376514
Garibay, O. O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., ... Xu, W. (2023). Six Human-Centered Artificial Intelligence Grand Challenges. *International Journal of Human–Computer Interaction*, *39*(3), 391–437. Retrieved from https://doi.org/10.1080/10447318.2022.2153320

Ghajargar, M., & Bardzell, J. (n.d.). Making AI Understandable by Making It Tangible: Exploring the Design Space with Ten Concept Cards. In *Proceedings of the 34th Australian Conference on Human-Computer Interaction* (pp. 74–80). Association for Computing Machinery. Retrieved 2023-04-26, from `https://dl.acm.org/doi/10.1145/3572921.3572942`

Ghajargar, M., Bardzell, J., Smith-Renner, A. M., Höök, K., & Krogh, P. G. (n.d.). Graspable AI: Physical Forms as Explanation Modality for Explainable AI. In *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 1–4). Association for Computing Machinery. Retrieved 2023-04-26, from `https://doi.org/10.1145/3490149.3503666`

Gillick, J., & Bamman, D. (2021). What to Play and How to Play it: Guiding Generative Music Models with Multiple Demonstrations. In *NIME 2021.*

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Pedreschi, D., & Giannotti, F. (2018). *A Survey Of Methods For Explaining Black Box Models.*

Gunning, D. (2016, Aug). Explainable Artificial Intelligence (XAI). *DARPA/I2O Proposers Day.*

Hadjeres, G., Nielsen, F., & Pachet, F. (2017). *GLSR-VAE: Geodesic Latent Space Regularization for Variational AutoEncoder Architectures.*

Hazzard, A., Greenhalgh, C., Kallionpää, M., Benford, S., Veinberg, A., Kanga, Z., & McPherson, A. (2019). Failing with Style: Designing for Aesthetic Failure in Interactive Performance. In *2019 CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019) Glasgow, Scotland, UK.* (p. 13 pages). ACM, New York, NY, USA.

Healey, P. G. T., Leach, J., & Bryan-Kinns, N. (2005). Inter-Play : Understanding Group Music Improvisation as a Form of Everyday Interaction..

Heer, M. C., Kale, A., & Jeffrey. (2019). Capture & Analysis of Active Reading Behaviors for Interactive Articles on the Web. *Computer Graphics Forum (Proc. EuroVis).*

Hermann, T., Hunt, A., & Neuhoff, J. G. (2011). *The sonification handbook* (Vol. 1). Logos Verlag Berlin.

Hermann, T., & Reinsch, D. (2021). Sc3nb: A Python-SuperCollider Interface for Auditory Data Science. In *Audio Mostly 2021* (pp. 208–215).

Herremans, D., Chuan, C., & Chew, E. (2017, September). A Functional Taxonomy of Music Generation Systems. *ACM Computing Surveys*, *50*(5). Retrieved from `https://doi.org/10.1145/3108242`

Hiller, L., & Isaacson, L. (1957). *Illiac Suite: For String Quartet.* New Music Edition.

Hoffman, G., & Weinberg, G. (2010). Shimon: an interactive improvisational robotic marimba player. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems* (pp. 3097–3102).

Huang, Q., Park, D. S., Wang, T., Denk, T. I., Ly, A., Chen, N., ... Han, W. (2023). *Noise2music: Text-conditioned music generation with diffusion models.*

Jourdan, T., & Caramiaux, B. (2023a). Culture and Politics of Machine Learning in NIME: A Preliminary Qualitative Inquiry. In *New Interfaces for Musical Expression.*

Jourdan, T., & Caramiaux, B. (2023b). Machine Learning for Musical Expression: A Systematic Literature Review. In *New Interfaces for Musical Expression.*

Kandinsky, W. (1912). *Concerning the Spiritual in Art* (M. T. H. Sadler, Trans.). Dover Publications Inc. (1972 translation).

Karimi, P., Rezwana, J., Siddiqui, S., Maher, M. L., & Dehbozorgi, N. (2020). Creative sketching partner: An analysis of human-ai co-creativity. In *Proceedings of the 25th international conference on intelligent user interfaces* (p. 221–230). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3377325.3377522`

Kawai, L., Esling, P., & Harada, T. (2020). Attributes-aware deep music transformation. In *Proceedings of the international society for music information retrieval conference (ismir).*

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kozierok, R., Aberdeen, J., Clark, C., Garay, C., Goodman, B., Korves, T., ... Peterson,

M. W. (2021). Hallmarks of Human-Machine Collaboration: A framework for assessment in the DARPA Communicating with Computers Program.

Krajna, A., Kovac, M., Brcic, M., & Šarčević, A. (2022, May). Explainable Artificial Intelligence: An Updated Perspective. In *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)* (pp. 859–864).

Kramer, G. (1993). *Auditory display: Sonification, audification, and auditory interfaces.* Perseus Publishing.

Lample, G., Zeghidour, N., Usunier, N., Bordes, A., Denoyer, L., & Ranzato, M. (2018). *Fader Networks: Manipulating Images by Sliding Attributes.*

Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (p. 1–15). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3313831.3376590`

Liu, Z., Michaud, E. J., & Tegmark, M. (2023, March). *Omnigrok: Grokking Beyond Algorithmic Data* (No. arXiv:2210.01117). arXiv.

Long, D., Padiyath, A., Teachey, A., & Magerko, B. (2021a). The Role of Collaboration, Creativity, and Embodiment in AI Learning Experiences. In *Creativity and Cognition.* New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3450741.3465264`

Long, D., Padiyath, A., Teachey, A., & Magerko, B. (2021b). The Role of Collaboration, Creativity, and Embodiment in AI Learning Experiences. In *In Creativity and Cognition (C&C '21), June 22, 23 Virtual Event, Italy.* (pp. 1–10). ACM, NY, USA,. Retrieved from `https://doi.org/10.1145/3450741.3465264`

Louie, R., Coenen, A., Huang, C. A., Terry, M., & Cai, C. J. (2020a). Cococo: AI-Steering Tools for Music Novices Co-Creating with Generative Models. In *IUI '20 Workshops, March 17, 2020, Cagliari, Italy.* (pp. 1–6). ACM, NY, USA.

Louie, R., Coenen, A., Huang, C. Z., Terry, M., & Cai, C. J. (2020b). Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (p. 1–13). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3313831.3376739`

Louie, R., Cohen, A., Huang, C. A., Terry, M., & Cai, C. J. (2020). Cococo: AI-Steering Tools for Music Novices Co-Creating with Generative Models. In *HAI-GEN+ user2agent@ IUI.*

Lovelace, A. A. (1843). Notes by August Ada Lovelace. *Taylor's Scientific Memoirs*, *III*, 666–731.

Lubart, T. (2005). How can computers be partners in the creative process: Classification and commentary on the Special Issue. *International Journal of Human-Computer Studies*, *63*(4), 365–369. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1071581905000418` (Computer support for creativity)

Lupker, J. A. T. (2021). Score-Transformer: A Deep Learning Aid for Music Composition. In *NIME 2021.*

Lyu, Z., Li, J., & Wang, B. (2021, nov). AIive: Interactive visualization and sonification of neural networks in virtual reality. In *2021 IEEE international conference on artificial intelligence and virtual reality (AIVR).* IEEE.

Malsattar, N., Kihara, T., & Giaccardi, E. (2019). Designing and Prototyping from the Perspective of AI in the Wild. In *Proceedings of the on Designing Interactive Systems Conference (DIS '19), June 23–28, San Diego, CA, USA.* Retrieved from `http://dx.doi.org/10.1145/3322276.3322351`

Manco, I., Benetos, E., Quinton, E., & Fazekas, G. (2021). *Muscaps: Generating captions for music audio.*

McCormack, J., Gifford, T., Hutchings, P., Llano Rodriguez, M. T., Yee-King, M., & d'Inverno, M. (2019). In a Silent Way: Communication Between AI and Improvising Musicians Beyond Sound. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (p. 1–11). New York, NY, USA: Association for Computing Machinery. Retrieved from

https://doi.org/10.1145/3290605.3300268
Morabito, R., Armitage, J., & Magnusson, T. (2022, June). Ritualistic approach to sonic interaction design: A poetic framework for participatory sonification. Georgia Institute of Technology.
Morreale, F., & McPherson, A. (2017, 05). Design for Longevity: Ongoing Use of Instruments from NIME 2010-14. In *Proc. NIME'17, May 15-19, 2017, Aalborg University Copenhagen, Denmark* (pp. 192–197).
Murray-Browne, T., & Tigas, P. (2021a). Emergent interfaces: Vague, complex, bespoke and embodied interaction between humans and computers. *Applied Sciences*, *11*(18). Retrieved from https://www.mdpi.com/2076-3417/11/18/8531
Murray-Browne, T., & Tigas, P. (2021b, 4 29). Latent Mappings: Generating Open-Ended Expressive Mappings Using Variational Autoencoders.. Retrieved from https://nime.pubpub.org/pub/latent-mappings (https://nime.pubpub.org/pub/latent-mappings)
Nimkulrat, N. (2020, September). Translational craft: Handmade and gestural knowledge in analogue–digital material practice. *Craft Research*, *11*(2), 237–260.
Noel-Hirst, A., & Bryan-Kinns, N. (2023). An autoethnographic exploration of xai in algorithmic composition. In *Explainable ai for the arts workshop at acm creativity and cognition 2023.*
Nyrup, R., & Robinson, D. (2022, February). Explanatory Pragmatism: A Context-Sensitive Framework for Explainable Medical AI. *Ethics and Information Technology*, *24*(1), 13.
OpenAI. (2022). *Chatgpt.* Retrieved from https://chat.openai.com/
Pachet, F. (2003). The Continuator: Musical Interaction With Style. *Journal of New Music Research*, *32*(3), 333–341. Retrieved from https://www.tandfonline.com/action/journalInformation?journalCode=nnmr20
Papetti, S., & Saitis, C. (2018). *Musical haptics.* Springer Nature.
Pati, A., & Lerch, A. (2019, 06). Latent Space Regularization for Explicit Control of Musical Attributes..
Pati, A., & Lerch, A. (2020). *Attribute-based Regularization of Latent Spaces for Variational Auto-Encoders.*
Pati, A., Lerch, A., & Hadjeres, G. (2019). Learning to traverse latent spaces for musical score inpainting. *arXiv preprint arXiv:1907.01164*.
Payne, C. (2019). *Musenet.* https://openai.com/blog/musenet/.
Pelinski, T., Shepardson, V., Symons, S., Caspe, F., Temprano, A., Armitage, J., ... McPherson, A. (2022, June 28). Embedded ai for nime: Challenges and opportunities. In *Nime.*
Pessoa, J., & Deloumeaux, L. (2009). *The 2009 UNESCO Framework for Cultural Statistics (FCS).* United Nations Educational, Scientific and Cultural Organization.
Proctor, R., & Martin, C. P. (2020). A Laptop Ensemble Performance System using Recurrent Neural Networks. Proc. New Interfaces for Musical Expression (NIME'20).
Quellec, G., Al Hajj, H., Lamard, M., Conze, P., Massin, P., & Cochener, B. (2021). ExplAIn: Explanatory artificial intelligence for diabetic retinopathy diagnosis. *Medical Image Analysis*, 102118.
Rau, S., Heyen, F., Wagner, S., & Sedlmair, M. (2022, December). Visualization for ai-assisted composing. In *Proceedings of the 23rd International Society for Music Information Retrieval Conference* (p. 151-159). Bengaluru, India: ISMIR. Retrieved from https://doi.org/10.5281/zenodo.7316618
Reed, C. N., & McPherson, A. P. (2023). The Body as Sound: Unpacking Vocal Embodiment through Auditory Biofeedback. In *Proceedings of the Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3569009.3572738
Roberts, A., Engel, J., Mann, Y., Gillick, J., Kayacik, C., Nørly, S., ... Eck, D. (2019). Magenta Studio: Augmenting Creativity with Deep Learning in Ableton Live. In *Proceedings of the International Workshop on Musical Metacreation (MUME).* Retrieved from http://musicalmetacreation.org/buddydrive/file/mume_2019_paper_2/
Sajja, S., Aggarwal, N., Mukherjee, S., Manglik, K., Dwivedi, S., & Raykar, V. (2021). Ex-

plainable AI based Interventions for Pre-season Decision Making in Fashion Retail. In *8th ACM IKDD CODS and 26th COMAD* (pp. 281–289).

Sanchez, T., Caramiaux, B., Thiel, P., & Mackay, W. E. (2022, March). Deep Learning Uncertainty in Machine Teaching. In *IUI 2022 - 27th Annual Conference on Intelligent User Interfaces.*

Sarmento, P., Kumar, A., Chen, Y.-H., Carr, C., Zukowski, Z., & Barthet, M. (2023). Gtr-ctrl: Instrument and genre conditioning for guitar-focused music generation with transformers. In C. Johnson, N. Rodríguez-Fernández, & S. M. Rebelo (Eds.), *Artificial intelligence in music, sound, art and design* (pp. 260–275). Cham: Springer Nature Switzerland.

Schneider, F., Jin, Z., & Schölkopf, B. (2023). *Moûsai: Text-to-music generation with long-context latent diffusion.*

Shen, Y., Jiang, S., Chen, Y., Yang, E., Jin, X., Fan, Y., & Campbell, K. D. (2020). To explain or not to explain: A study on the necessity of explanations for autonomous vehicles. *arXiv preprint arXiv:2006.11684*.

Shneiderman, B. (2022). *Human-centered ai.* Oxford University Press.

Shneiderman, B., Fischer, G., Czerwinski, M., Resnick, M., Myers, B., Candy, L., ... Terry, M. (2006). Creativity Support Tools: Report From a U.S. National Science Foundation Sponsored Workshop. *International Journal of Human-Computer Interaction*, *20*(2), 61–77. Retrieved from `https://doi.org/10.1207/s15327590ijhc2002{_}1`

Steinbeck, P. (2022). 5 George Lewis, Voyager. In *Sound Experiments: The Music of the AACM* (pp. 109–120). Chicago: University of Chicago Press. Retrieved from `https://doi.org/10.7208/chicago/9780226820439-007`

Sturm, B. (2022). Generative ai helps one express things for which they may not have expressions (yet). In *Workshop on generative ai and hci at the chi conference on human factors in computing systems 2022.*

Sturm, B. L., & Ben-Tal, O. (2017, September). Taking the Models back to Music Practice: Evaluating Generative Transcription Models built using Deep Learning. *Journal of Creative Music Systems*, *2*(1). Retrieved from `https://doi.org/10.5920/jcms.2017.09`

Sturm, B. L., Ben-Tal, O., Monaghan, Ú., Collins, N., Herremans, D., Chew, E., ... Pachet, F. (2019). Machine learning research that matters for music creation: A case study. *Journal of New Music Research*, *48*(1), 36–55.

Sturm, B. L., Ben-Tal, O., Úna Monaghan, Collins, N., Herremans, D., Chew, E., ... Pachet, F. (2019). Machine learning research that matters for music creation: A case study. *Journal of New Music Research*, *48*(1), 36–55. Retrieved from `https://doi.org/10.1080/09298215.2018.1515233`

Sturm, B. L., Santos, J. F., Ben-Tal, O., & Korshunova, I. (2016). Music transcription modelling and composition using deep learning. *ArXiv*, *abs/1604.08723*.

Sturm, B. L., Santos, J. F., & Korshunova, I. (2015). Folk Music Style Modelling by Recurrent Neural Networks with Long Short Term Memory Units. *Extended Abstracts for the Late-Breaking Demo Session of the 16th International Society for Music Information Retrieval Conference*.

Suh, M. M., Youngblom, E., Terry, M., & Cai, C. J. (2021). AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3411764.3445219`

Tan, H. H., & Herremans, D. (2020). *Music FaderNets: Controllable Music Generation Based On High-Level Features via Low-Level Feature Modelling.*

Thelle, N. J. W., & Pasquier, P. (2021, 4 29). Spire Muse: A Virtual Musical Partner for Creative Brainstorming.. Retrieved from `https://nime.pubpub.org/pub/wcj8sjee` (https://nime.pubpub.org/pub/wcj8sjee)

Toussaint, G. (2002). A mathematical analysis of African, Brazilian, and Cuban clave rhythms. In *Townson University* (pp. 157–168).

Vigliensoni, G., McCallum, L., & Fiebrink, R. (2020, June). Creating Latent Spaces for

Modern Music Genre Rhythms Using Minimal Training Data. In *International Conference on Computational Creativity (ICCC).* Retrieved from `https://research.gold.ac.uk/id/eprint/29044/`

Wakkary, R. (2021). *Things We Could Design: For More Than Human-Centered Worlds* (K. Friedman & E. Stolterman, Eds.). Cambridge, MA, USA: MIT Press.

Wallis, I., Ingalls, T., Campana, E., & Vuong, C. (2013). Amateur Musicians, Long-term Engagement, and HCI. In *Music and Human-Computer Interaction* (pp. 49–66). Springer.

Wang, H. (2023). *Diffuseroll: Multi-track multi-category music generation based on diffusion model.*

Weger, M., Hermann, T., & Höldrich, R. (2018). Plausible Auditory Augmentation of Physical Interaction. , 8.

Weger, M., Hermann, T., & Höldrich, R. (2022). AltAR/table: A platform for plausible auditory augmentation. In *ICAD. Virtual Conference.*

Whorley, R., & Laney, R. (2020). Generating subjects for pieces in the style of bach's two-part inventions. In B. Sturm & A. Elmsley (Eds.), *The 2020 joint conference on ai music creativity.* Retrieved from `http://oro.open.ac.uk/72735/`

Xambó, A. (2022). Virtual Agents in Live Coding: A Review of past, present and future directions. *e-Contact!*, *21*(1).

Zhang, Y., Xia, G., Levy, M., & Dixon, S. (2021). COSMIC: A Conversational Anterface for Human-AI Music Co-Creation. In *NIME 2021.*

Zhu, J., Liapis, A., Risi, S., Bidarra, R., & Youngblood, G. M. (2018). Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation. *IEEE Conference on Computatonal Intelligence and Games, CIG*, *2018-August*.

Zukowski, Z., & Carr, C. (2018). Generating black metal and math rock: Beyond bach, beethoven, and beatles. *arXiv preprint arXiv:1811.06639*.