A Short Review of Responsible AI Music Generation

Elizabeth Wilson
Creative Computing Institute
University of the Arts London
London, UK
e.j.wilson@arts.ac.uk

Anna Wszeborowska
Creative Computing Institute
University of the Arts London
London, UK
a.wszeborowska@arts.ac.uk

Nick Bryan-Kinns
Creative Computing Institute
University of the Arts London
London, UK
n.bryankinns@arts.ac.uk

Abstract

Artificial intelligence continues to become increasingly embedded in musical practice and yet there is little evaluation of how transparent and ethical these systems are. Surveys of AI models to date focus on the technical features of AI models and there is a lack of surveys of the practical and ethical application of AI models for musicians, producers, and composers. This paper surveys 27 contemporary AI models for music generation in terms of creative input and output (symbolic music, audio, text, or image), musical task (symbolic composition or audio generation), and Responsible AI properties of transparency & explainability, fairness, accountability, and ethical AI. Analysis of these facets of AI model use in creative practice highlights the trade-offs and challenges in designing equitable and ethical AI tools for music-making. The survey highlights a lack of transparency and control of AI model training and fine-tuning, a lack of openness of licensing and source code, a lack of ethical reporting of training datasets, and a focus on AI models for audio generation at the expense of real-time music generation for use in composition, performance, and improvisation. Our analysis offers insights for researchers, developers, and musicians seeking to navigate this fast evolving landscape of musical AI. We suggest that research is needed to develop clearer frameworks for evaluating AI models in creative domains, focussing especially on user journeys that help users understand the mechanics, limitations, and ethical considerations of these systems in music making practice.

1 Introduction

The rapid advancement of generative AI (GenAI) models for music has introduced a vast array of tools that assist and automate aspects of musical creation, ranging from composition and sound design to text-based music generation. At the same time there has been increasing concern about how these AI models can be used ethically and responsibly. Whilst extensive surveys of AI models for music have been undertaken e.g. Herremans et al. (2018) and Xambó (2021), these focus on the technical features of AI models and there is a lack of surveys of the practical, responsible, and ethical application of AI models for musicians, producers, and composers. Garibay et al. (2023) defines the *responsible design of AI* as one of six "Human-Centered Artificial Intelligence Grand Challenges". Their definition emphasises designing for legal, ethical, and moral considerations in AI applications by systematically adopting Responsible AI principles of: *transparency & explainability*, *fairness*,

Proceedings of the 6th Conference on AI Music Creativity (AIMC 2025),
Brussels, Belgium, September 10th-12th

accountability, and *ethical AI* (ibid.). This paper offers an initial survey of popular AI models for generating music through the lens of Responsible AI principles.

Moreover, a critical reflection on the surveyed models reveals a fundamental issue: many AI music tools do not actively facilitate artistic interaction beyond simple input-output mechanisms. While some systems allow for real-time interaction or iterative refinement of music, a significant portion of AI-generated music emerges as fully formed, offering little room for user intervention beyond broad stylistic choices. This trend shifts the role of the musician from that of a composer to a curator, raising questions about creative agency in AI-assisted music-making, much as has been found in other creative practices where a shift is noted from "material production to critical integraton" (Sarkar, 2023, p.1).

2 Background

The increased integration of generative AI into music-making has given rise to complex questions surrounding ethical responsibility and creative authorship. These concerns also intersect with broader cultural and technological shifts, prompting discussion on who—or what—can be considered a legitimate creative agent in any artistic process. Furthermore, these tools and workflows offer musicians new possibilities for composition, performance, and collaboration, but they also introduce challenges related to bias, transparency, and the evolving relationship between human and machine creativity, to name a few.

AI systems rely heavily on human involvement, particularly in the case of Machine Learning (ML)—one of the most prevalent approaches in AI development today (Kashyap and Kumar, 2019). ML models are trained on large datasets, often guided by predefined "correct" metrics and answers—an inherently subjective notion in artistic contexts. As AI becomes progressively further embedded in creative workflows, human involvement should extend beyond dataset creation to include key decisions around model training, curation, and interpretation.

As such, the problem space for Ethical and Responsibile AI and the Arts is vast, encompassing a multitude of challenges ranging from technical considerations to societal impacts. This paper does not aim to exhaustively address every dimension of these intersecting issues. For more comprehensive explanations of broader questions – such as those concerning attribution and authorship, legal accountability, privacy, trust, and governance – readers are referred to Garibay et al. (2023); Morreale (2021); Piskopani et al. (2023); Newman et al. (2023); Hagendorff (2024).

Generative AI plays a large role in current music and AI practice. Increasingly, these generative systems are producing outputs that are passable, or even unrecognisably different to human-like content (Oppenlaender, 2022; Yang et al., 2023; Feuerriegel et al., 2024). However, many commercially available GenAI tools require huge training datasets. A significant amount of effort has been focused on scaling up dataset sizes for training models so that these systems can produce extremely high quality samples (e.g. (Chen and Du, 2021; Huang et al., 2023)). Diffusion models are a current advancement in the generation process, many of which contain these scaled up datasets. For example, Song and Ermon (2019); Ho et al. (2020) describe generative models that generate samples by iteratively de-noising random noise. In general, diffusion models have shown the capability to generate high quality images (Ho et al., 2022b) and video (Ho et al., 2022a) and have also been applied to neural audio synthesis (Yang et al., 2023). However, their application to audio remains relatively inefficient due to inconsistencies between generated samples, which can lead to artifacts, temporal instability, and challenges in achieving coherent long-term structure.

Despite these advancements, the increasing complexity and opacity of generative AI models raise pressing concerns regarding transparency, accountability, and creative agency. The reliance on large-scale datasets—often scraped from the internet without clear consent or attribution—compounds issues related to bias, ownership, and ethical usage (Morreale et al., 2024). Unlike traditional artistic tools, which serve as extensions of human intent, many contemporary AI music systems function as opaque generators, producing outputs that users have limited control over or understanding of. This lack of interpretability not only challenges the role of the artist in AI-assisted workflows but also restricts meaningful human-AI collaboration by reducing creative decision-making to high-level prompt engineering or parameter tweaking (Bryan-Kinns, 2024).

Name	Year	Website	Input Type	Output Type	Output Length	Real Time	AI Technique	License	Free	Open Source	Checkpoints	Fine-tune	Train From Scratch	Training Dataset
Aimi	2023	aimi.fm	None/Text	Audio	Endless	Yes	N/S	Proprietary	No	No	No	No	No	Not disclosed
AIVA	2019	aiva.ai	MIDI/Audio	Audio/MIDI	5min30s	No	N/S	Proprietary	Yes/No	No	No	No	No	Not disclosed
Amadeus Code	2019	amadeuscode.com	Genre	Audio/MIDI	Variable	No	N/S	Proprietary	Yes/No	No	No	No	No	Not disclosed
Audialab	2021	audialab.com	Text	Audio	Short	Yes	N/S	Proprietary	No	No	No	No	No	Not disclosed
AudioLCM	2023	github.com/Text-to-Audio/AudioLCM	Text	Audio	Variable	Yes	Latent Consistency Model (LCM)	MIT	Yes	Yes	Yes	Yes		"Teacher" model not disclosed, AudioCaps dataset (Kim et al., 2019) for AudioLCM model
Boomy	2019	boomy.com	Text/Audio	Audio	~2 mins/track	No	N/S	Proprietary	Yes (Paid features)	No	No	No	No	Not disclosed
Dance Diffusion	2022	github.com/harmonai-org/sample-generator	Text/Audio	Audio	Variable	No	Latent Diffusion	MIT	Yes	Yes	Yes	Yes	Yes	glitch.cool, songaday.world, MAESTRO, Unlocked Recordings, xeno-canto.org
Loudly	2023	loudly.com	Text/Metadata	Audio	7min	No	N/S	Proprietary	Yes/No	No	No	No	No	Not disclosed
Magenta Continue	2018	magenta.tensorflow.org/studio#continue	MIDI	MIDI	32 Bars	No	LSTM	Apache 2.0	Yes	Yes	Yes	Yes	Yes	MelodyRNN: Not disclosed; PerformanceRNN: The Piano-e-Competition dataset
Magenta Drumify	2018	magenta.tensorflow.org/studio#drumify	MIDI	MIDI	Input Length	No	LSTM	Apache 2.0	Yes	Yes	Yes	Yes	Yes	Expanded Groove MIDI, Groove MIDI
Magenta Generate	2018	magenta.tensorflow.org	None	MIDI	4 Bars	No	VAE	Apache 2.0	Yes	Yes	Yes	Yes	Yes	NSynth, MAESTRO, Lakh MIDI
MAGNet	2019	github.com/Louismac/MAGNet	Audio	Audio	Input Length	Yes	LSTM	BSD 3-Clause	Yes	Yes	No	No	Yes	Not disclosed
Mubert	2016	mubert.com	Text/Image	Audio	25min	No	N/S	Proprietary	Yes/No	No	No	~	No	Not disclosed
MusicGen	2023	ai.honu.io/papers/musicgen	Text	Audio	30s	No	Transformer	MIT/CC-BY-NC	Yes	Yes	10 Models	Yes	Yes	NSynth Dataset; Others not disclosed
MusicLM	2023	google-research.github.io/seanet/musiclm/examples/	Text	Audio	30s	No	Transformer	Proprietary	Yes	No	No	No	No	MusicCaps; AudioSet
Okio Nendo	2023	okio.ai/	Audio/Text	Audio	Variable	Yes	Suite of AI tools	MIT for core tools	Yes	Yes	No	No	No	Not disclosed
RAVE	2022	acids-ircam.github.io/RAVE/	Audio	Audio	Variable	Yes	VAE	MIT	Yes	Yes	Yes	Yes	Yes	Can train on custom datasets, e.g. IILGuitarTimbre, CSTR VCTK Corpus
Riffusion	2022	riffusion.com/	Text	Audio	3min	No	Diffusion	MIT	Yes	Yes	Yes	~	No	Not disclosed
R-VAE	2022	github.com/vigliensoni/R-VAE-models	MIDI	MIDI	2 Bars	Near	VAE	GPLv3	Yes	Yes	No	No	No	The Future Sample Pack
SampleRNN	2016	github.com/soroushmehr/sampleRNN_ICLR2017	Audio	Audio	Variable	No	Hierarchical Recurrent Neural Network (RNN)	MIT	Yes	Yes	Community-based	~	Yes	Not disclosed
Soundful	2023	soundful.com/	Text/Metadata	Audio/MIDI	2min30s	No	N/S	Proprietary	Yes/No	No	No	No	No	Not specified
Soundraw	2024	soundraw.io/	Metadata	Audio	5m	No	N/S	Proprietary	Yes/No	No	No	No	No	Not disclosed
Splash	2023	splashmusic.com/	Text	Audio	3m	No	N/S	Proprietary	Yes/No	No	No	~	No	Not disclosed
Stable Audio	2023	stability.ai/news/introducing-stable-audio-open	Text/Audio	Audio	3min	No	Diffusion	~	Yes	Yes	Yes	Yes	Yes	freesound.org; freemusicarchive.org
Suno	2023	suno.com/	Text	Audio	2min	No	N.S.	Proprietary	Yes/No	No	*	Yes	No	Not disclosed
Udio	2024	udio.com/	Text	Audio	30s	No	N.S.	Proprietary	Yes	No	No	*	No	Not disclosed
WaveNet	2016	deepmind.google/discover/blog/wavenet-a-generative-model-for-raw-audio/	Audio	Audio	Variable	No	Autoregressive Convolutional Neural Network	Proprietary (multiple open-source reimplementations exist)	No (Yes OS Vs)	No (Yes OS Vs)	Community-based	Yes	Yes	Not disclosed

Table 1: Generative AI Systems Surveyed Comparing Various Facets of Responsible AI Practices

3 Survey of AI Music Models

This review examines 27 contemporary GenAI music tools and models. These were selected for this review to be representative of the current ML landscape of GenAI for music. The selection process was guided by the goal of capturing a broad range of approaches to AI music generation, in both academic and commercial domains. Specifically, we prioritised tools that have been publicly available or documented since 2016, a period marked by the rise of deep learning-based generative methods that significantly advanced the field. Systems were included to reflect a range of characteristics—from proprietary, closed-source platforms to fully open-source frameworks. Particular attention was given to tools that have been used in actual music-making contexts or discussed in critical or industry discourse.

Each tool was categorised in terms of several key parameters: *input* (text, symbolic music, audio, image, or metadata) and *output* (symbolic music, or audio), *output length*, whether it operates in *real-time* or not, what AI technique is used for generation, what *licences* are needed, whether it is *free* and *open-source*, and finally features of the model itself including whether it offers *checkpoints*, *fine-tuning*, the ability to *train from scratch*, and what *training datasets* are used (see Table 1). We then reflect on these AI models in terms of how they support music making tasks and how they relate to Responsible AI principles of transparency & explainability, fairness, accountability, and ethical AI. By doing so, we aim to provide a clearer understanding of how AI is currently used in music-making and the implications of these technologies for creative practice and responsible AI design.

The data from the review has been transformed into an online, interactive tool, as described in Bryan-Kinns et al. (2025b). This tool allows users to explore models by using the table headings as searchable fields. It is also open-source, encouraging users to contribute by adding new models to the database.

This review focuses solely on AI-driven generative and interactive music systems, and does not cover areas such as AI-powered audio effects and mixing, which include tools for automatic mastering, adaptive equalisation, and intelligent reverb design. Though these technologies play a significant role in the process of music interface design, and are still interesting to evaluate from the standpoint of creative agency, they primarily function as enhancements to existing workflows rather than acting to generate agents in composition or performance.

Similarly, this paper does not delve into analytical and computational tools, such as AI-driven musicology, style recognition, or recommendation systems. While these systems contribute valuable insights into musical structure and influence creative decisions, they operate more as facilitators of discovery rather than as direct participants in the act of music-making. By narrowing our scope to models that engage directly with creative agency, we aim to provide an exploration of the ways in which AI entangles with human artistic expression. Furthermore, this paper does not provide any aesthetic or technical evaluations of systems and their limitations. Whilst this is an important facet of understanding how human-machine interaction can be facilitated, it is outside the scope of this review.

4 Music Generation

Generative AI models for music composition refer to models that create various elements of musical structure, such as melody, harmony, chords, and rhythm. These models can operate in either a symbolic format (most commonly MIDI) or directly generate audio-based compositions - i.e. the *output* of the model. Symbolic models focus on structured representations of music that can be further arranged, edited, and performed using digital audio workstations (DAWs) or notation software; whereas audio-based compositions give less flexibility in terms of post-generation modification. While symbolic models allow users to manipulate individual musical elements—such as altering a melody, re-harmonising chords, or adjusting rhythmic patterns—audio-based models generate complete sound outputs that are more difficult to deconstruct and edit at a granular level.

4.1 Symbolic music generation

We found that 7 of the 27 AI models surveyed were capable of symbolic music generation compared to 20 of the 27 that generated audio (*output*). Models that produce symbolic content are useful for musicians to explore new musical ideas, generate variations or extend creative ideas. They are often used in conjunction with common DAWs, and plug-ins for DAWs such as Ableton Live have made these more accessible to practitioners.

One of the earliest contributions to symbolic music generation comes from Google's Magenta project, which released several models in 2018. Magenta Continue extends pre-existing musical sequences, using the predictive power of recurrent neural networks (RNN) to generate notes that are likely to follow your drum beat or melody, providing AI-assisted melodic continuation and variation (Google, 2018a). Magenta Generate employs a hierarchical latent vector model to produce structured long-form compositions with musical coherence (Google, 2018c). Magenta Interpolate, unlike the previous systems, takes two drum beats or two melodies as inputs. It then generates up to 16 clips which combine the qualities of these two clips. Another notable model from the same initiative, Magenta Drumify, transforms melodic inputs into rhythmic accompaniments, adding depth and dynamism to musical arrangements (Google, 2018b).

In 2019, the AIVA-model was released (AIVA, 2019), described as a "music generation assistant that allows you to generate new songs in more than 250 styles". They also advertise that users can: "Create your own style models. Upload an audio or MIDI influence. Edit your generated tracks. Download in any file format." The interface of the AIVA system is similar to a DAW, where users can generate different tracks, and edit and recompose at the MIDI note level using the software.

In contrast, R-VAE / RhythmVAE approached the challenge of symbolic music generation by introducing a variational autoencoder-based approach to rhythmic pattern generation, leveraging latent representations of rhythm to produce novel drum sequences that align with diverse musical contexts (Vigliensoni et al., 2022). In the creation of this model, the authors found the non-linearities of the learned latent spaces coupled with tactile interfaces to interact with the models were very expressive and led to unexpected places in musical composition and live performance settings requiring *real-time* generation.

4.2 Audio generation

Unlike symbolic models, which produce structured representations (such as, but not limited to, MIDI) that can be edited and arranged, audio-based composition models generate fully rendered music in an audio format. These models often prioritise automation, enabling users to create music with minimal technical expertise. However, the shift from symbolic to audio-based generation also reduces user control over musical elements, raising questions about what support there is for creative agency and authorship.

Amadeus Code (Amadeus, 2019) was one of the earlier AI systems designed to generate melodies and harmonies in both audio and MIDI formats, allowing composers to integrate AI-generated ideas into their workflows. Around the same time, Boomy (BOOMY, 2019) emerged as a platform enabling users to create entire tracks by selecting a style or genre, streamlining the music production process for non-musicians. More recent tools such as Soundful (Soundful, 2023), Loudly (Loudly, 2023), and SoundRaw (Soundraw, 2024) further refine AI-assisted composition by providing automated music generation tailored for content creators, background music production, and commercial applications.

One of the most prolific models in this area is Suno (Suno, 2023), which allows users to generate music from text descriptions, offering a flexible and intuitive platform for music creation. Splash (Splash, 2023) follows suit, providing users with the ability to generate music based on text prompts *inputs*, with additional features for customisation. Udio (Udio, 2024), which is still currently in its beta phase (ibid.), is another model in the text-to-music field, facilitating music composition based on user-defined text prompts.

Further expanding on this model, Stable Audio (Evans et al., 2024) provides a new approach to text-to-music generation, using a latent diffusion model to create coherent musical compositions from textual inputs. Similarly, Dance Diffusion (Harmonai-org, 2022) allows for the creation of dance-oriented music through textual prompts, focusing on rhythmic and stylistic elements that align with contemporary electronic music trends. MusicGen (Copet et al., 2024) and MusicLM (Agostinelli et al., 2023) both provide advancements in text-to-music generation, with the former employing simple yet highly controllable models for music composition, while the latter provides high-quality, complex compositions through a deep learning framework. Another unique model, Riffusion (Riffusion, 2019), generates music from both text descriptions and sound samples, offering a hybrid approach to text-to-music generation. Finally, AudioLCM (Liu et al., 2024) builds on latent consistency models to generate music from text prompts, further advancing the ability of AI to create nuanced and stylistically rich compositions from written inputs.

Aimi (AIMI, 2024) is an interactive AI-assisted music composition platform that allows users to create music by selecting different inputs, such as genre and mood, effectively conditioning the music generation process. Mubert (Mubert, 2016) generates music not only from text prompts but also from images and specific stylistic cues, offering a diverse set of creative possibilities. These conditioned music generation models enable users to control and shape the music creation process, especially in settings where stylistic coherence and user input are key.

In the context of sound design, AI models such as MAGnet (McCallum, 2019) and Audialab (BELIBOU and IFTENE, 2024) generate audio samples and loops that can be used as building blocks in larger musical compositions. These tools allow users to create and manipulate sounds that would otherwise be time-consuming or technically challenging to produce, offering a more efficient way to explore new sonic textures and experiment with unconventional sounds. The latter is a web-based interface allowing users to create with a variety of models integrated into the platform, but no direct links to the source are provided. An ethical statement is provided on their website, to help in "differentiating responsible and irresponsible AI applications", alongside starting an organisation centred on the ethical use of such AI tools. However, there is no particular mission statement or constructive solutions to issues of agency, bias, or any other facet that could easily be present in these models.

Additionally, AI-driven sound generation models such as RAVE (Caillon and Esling, 2021) and SampleRNN (Mehri et al., 2016) focus on producing audio at the sample level, which can be used in a variety of musical genres or contexts. These models provide highly flexible tools for generating sounds that can mimic existing sonic textures or produce entirely new, unexpected results.

4.3 Real-time generation: musical composition, performance, and improvisation

Of the 27 AI models reviewed, only 6 support *real-time* interaction, while the remaining 21 focus on pre-rendered compositions (with one of these *close to* real-time). This distinction is particularly relevant in live performance and improvisational settings, where immediacy, responsiveness, and collaboration are critical.

Firstly, Aimi, is a platform that offers real-time composition by allowing users to steer generative processes by selecting musical parameters such as genre, mood, and intensity. They claim that "generation is very low latency and can deliver real-time results with sub 200 ms latency" (AIMI, 2024), which is "useful in the context of apps, games and other digital experiences delivering a scalable, adaptive audio solution that responds instantly to user input and in-app changes" (AIMI, 2024). However, live performance typically requires latency below 50ms (McPherson et al., 2016) making such systems unsuitable for performance in these contexts.

AudioLCM similarly supports interactive generation through their "consistency-based model tailored for efficient and high-quality text-to-audio generation" (Liu et al., 2024, p. 1). By leveraging Latent Consistency Models (LCMs), AudioLCM drastically reduces the number of denoising steps typically required by diffusion models. As a result, it achieves audio synthesis speeds up to 333× faster than real-time on a single NVIDIA RTX 4090 GPU (Liu et al., 2024), enabling responsive and practical deployment in interactive applications. This efficiency allows users to generate coherent, high-fidelity audio clips within seconds, which the authors claim makes it well-suited for real-time creative workflows such as live audio prototyping, music ideation, or adaptive sound design.

Thirdly MAGNet, developed at the Creative Computing Institute McCallum (2019), supports real-time through its integration with the Dorothy library for creative coding (McCallum, 2023). MAGNet employs a spectral approach to audio analysis and generation, wherein neural networks are trained to model time–frequency representations of audio signals. This design enables low-latency audio synthesis and transformation within interactive coding environments, supporting real-time experimentation and performance. The system's architecture emphasises responsiveness and creative flexibility, making it particularly suitable for live coding, generative composition, and performative sound design.

Similarly, RAVE is a variational autoencoder (VAE) model designed to achieve both fast and high-quality audio waveform synthesis (Caillon and Esling, 2021). It supports real-time generation through an optimised architecture capable of operating significantly faster than real-time, even on standard CPUs. The model has been integrated into a range of creative tools, including audio plug-ins and MaxMSP objects, thereby enabling musicians and sound designers to incorporate neural audio synthesis into their existing workflows. This real-time compatibility makes RAVE particularly well-suited for live performance, interactive composition, and exploratory sound design contexts (Wilson et al., 2023).

5 Responsible AI: Fairness

Generative systems often exhibit unfair behaviour where people are unintentionally treated differently. This may be due to bias in the training datasets or bias in the architecture of AI models themselves (Garibay et al., 2023). While music generation technologies open up creative possibilities, they also raise concerns about the training datasets used, the underlying algorithms, and the potential for biases in the generated music. Research demonstrates that GenAI architectures can bias output in addition to the bias of the datasets used to train the model (Bryan-Kinns et al., 2024). Without access to the training data or the methods used to train the models, users cannot fully assess the ethical implications of using AI-generated music. For example, the data used to train these models may reflect imbalances in the representation of certain genres, cultures, or sound aesthetics, which could result in the over-representation of specific styles and the under-representation of others. Only 5 of the 27 models declared which *training datasets* were used in training the AI model - this makes it very difficult for musicians to assess how fairly the AI models represent musical styles and cultures. Moreover, only 11 of 27 models declared which *AI technique*s are used to generate music. Slightly less than half of the AI models surveyed were available as *open-source* (11 of 27). This lack of openness limits opportunities for independent researchers and artists to scrutinize the potential biases embedded in the datasets or the fairness of the model's training process. Examples of good practice in this area include all of the Magenta Models being released Open Source on GitHub and as free Ableton plug-ins for musicians to interface with the generated MIDI.

Only 12 of 27 models could be *trained from scratch* by end-users. In other words, it is not possible for 15 of the models to be trained solely on musicians' original content and limits musicians to using AI models which may be unfairly biassed or under representative of diverse musical cultures. Fewer models (9) supported *fine-tuning* of pre-trained models to allow for customisation of the AI model.

Again, this prevents musicians from modifying AI models to address any bias or underrepresentation in the AI models.

6 Responsible AI: Ethics, Accountability and Liability

Responsible AI principles of ethics, accountability, and liability are intertwined when using Generative AI and music. Whilst RAI concerns about the accountability of AI models are less weighty when examining GenAI music systems than, say in life threatening situations such as autonomous vehicle accidents, there remain questions about who is liable for copyright and intellectual property infringements in the training and use of GenAI. Again, the lack of visibility of which *training datasets* were used to train the AI models leaves musicians open to uncertainty about whether an AI model was trained respecting copyright and intellectual property rights, and whether any generated music could later be considered in breach of copyright and intellectual property rights.

Licensing structures of the AI models surveyed often prioritise commercial interests, leaving creators with limited rights over the music they generate. For example, 12 of the models reviewed required proprietary *license*s for use. These factors contribute to an uneven distribution of creative control, where a few corporations hold considerable power over the music creation process, potentially stifling innovation and ethical considerations in the industry. For example, the AIVA model is not open-sourced, and varying user-pricing tiers exist to limit licensing practices based on a subscription model (AIVA, 2019), where free users must credit AIVA, and outputs cannot be monetised, whereas paid-subscribers can monetise outputs, and do not need to credit AIVA. This closed-source nature and tiered subscription model raise concerns from a responsible AI music perspective. The restricted access to its underlying algorithms limits transparency and reproducibility, while its licensing structure creates inequalities in creative ownership.

In contrast, R-VAE was released as a web interface tool and RhythmVAE, the model it was based on, could be exported as a Max4Live device, making it easy for musicians to integrate into their workflow. The model is Open-Source, free to interface with and Network Architecture discussed in length in the associated research paper (Vigliensoni et al., 2022). AudioLab model is open-source and released on Github, designed also to be used with the author's own open-source creative coding library. Both RAVE and SampleRNN are are examples of good practice in this area - open-sourced models with code openly available to share, code and fork, and datasets for training and implementing the model are exposed.

7 Responsible AI: Explainability and Transparency

Whilst explainability is widely recognised as underdeveloped in the Arts (Bryan-Kinns, 2024) including music (Bryan-Kinns et al., 2024), there are a small number of GenAI music systems that support transparency and control of the model itself.

Control over music generation is reduced in the audio-based AI models, where audio-based compositions emerge as fully formed outputs, often several minutes long (all *output length*s for audio generation are longer than for symbolic music generation), or even endless with Aimi. This leaves limited room for intervention beyond broad post-generation manipulations such as remixing or applying effects. Indeed, only 8 of the 27 models offer AI model *checkpoints* which allow users to better understand the training process undertaken and return to previous states of training and fine-tuning. In these cases, the user's role shifts from detailed compositional control to high-level curation, selecting and modifying AI-generated material rather than shaping its foundational structure. This shift in creative agency raises important questions for responsible AI practices – humans and machines collaborate more closely in creative processes, the balance of agency, authorship, and accountability becomes increasingly complex, highlighting the need for transparent and equitable frameworks that ensure both human creators and AI models have clearly defined roles in the creative process.

Text-prompting *input type* models make up 10 of the 27 models surveyed. These generate music based on written descriptions or verbal instructions, offer a means of artistic expression that is rooted in linguistic expression, but they also present significant challenges regarding creative agency, transparency, and bias. One of the primary concerns with text-prompting models is the distribution of creative agency. While these systems allow users to quickly generate music from textual descriptions, the process often minimises the role of the human artist in shaping the final output. In many cases, the creative control granted to users is limited to high-level prompts, such as genre selection or mood setting, while the underlying composition, arrangement, and orchestration are left to the AI with little transparency about the processes involved.

In terms of transparency, many text-prompting models operate as proprietary, black-box systems, where the algorithms and datasets driving the music generation are not made publicly available. This lack of transparency makes it difficult for users to fully understand how the AI produces music from their inputs, limiting their ability to assess the fairness or accuracy of the generated outputs. Without access to the inner workings of the model, it is challenging for users to detect potential biases in the data or the creative process.

8 Discussion

One of the persistent challenges in surveying current AI music systems is the lack of transparency around training data and model development practices. Indeed, a key challenge in conducting this survey was the difficulty of accessing information about features and Responsible AI practices of these models. Many companies and research groups do not disclose the datasets used to train their models, making it difficult to assess the provenance, representational scope, or potential copyright violations associated with the resulting outputs. In more concerning cases, such as those involving commercial platforms including Suno and Udio, developers have even acknowledged the potential use of copyrighted material in training data (Nayar, 2025). This admission raises critical legal and ethical questions around consent, fair use, and appropriation, especially when outputs closely mimic the structure or timbre of copyrighted recordings. Furthermore, many research papers and commercial platforms fail to provide clarity on aspects such bias mitigation strategies, and the extent of user control over AI-generated outputs. This lack of transparency itself raises ethical and creative concerns, particularly when AI-generated content is subject to proprietary licensing structures that limit artistic ownership and adaptability. As AI-generated music becomes increasingly prevalent, the need for clearer documentation of Responsible AI features including accountability, explainability and transparency, fairness, and ethics is more pressing than ever.

From the perspective of Responsible AI the use of AI-generated music, sound, and sample generation introduces the issue of authorship, as the creation of these outputs is often attributed to the algorithm rather than the user. This raises questions about the originality and ownership of AI-generated samples, especially when these sounds are incorporated into commercial or collaborative projects. Moreover, users may not have full control over the rights to the samples they generate, particularly when working with proprietary platforms that impose restrictive licensing agreements.

While these technologies open up creative possibilities, they also raise concerns about the transparency of the underlying algorithms and the potential for biases in the generated samples. Without access to the training data or the methods used to train the models, users cannot fully assess the ethical implications of using AI-generated samples. For example, the data used to train these models may reflect imbalances in the representation of certain genres, cultures, or sound aesthetics, which could result in the over-representation of specific styles and the under-representation of others.

Music making is a fundamental form of human creativity and yet in this survey we found that most of the AI systems reviewed offered only limited support for meaningful creative agency. For example, we found that many systems had low transparency around model behaviour, making it hard for musicians to understand what the AI was doing—a barrier to discovery and deeper creative engagement. Likewise, there was limited fine-grained control of music generation which again limits the creative freedom of musicians. It was also notable that many of the AI systems surveyed produced fully-composed songs leaving little room for iterative exploration, personalisation, or the reinsertion of the artist's distinctive voice beyond initial prompting.

Moreover, the lack of real-time interaction also limited the forms of creative practice to composition, and not supporting key forms of musical practice and expression such as improvisation and performance. We suggest that to promote human creative agency with GenAI models for music there needs to be more real-time interaction, more human-in-the-loop iteration of music being generated, and more transparency and control of the AI models themselves. In many ways these goals could be achieved by using smaller and more lightweight AI models which are able to operate in real-time on personal computers and whose workings can be more easily exposed and made understandable to musicians. In addition, smaller AI models would offer the chance to use smaller, potentially personally curated, datasets which would mitigate ethical concerns about the use of large training datasets and AI models outlined in this survey.

Overall, as AI-generated content becomes increasingly prevalent, the need for clearer documentation, especially within the scope of Responsible AI (Garibay et al., 2023), is more pressing than ever. Researchers, developers, and artists must play an active role in shaping their ethical and cultural trajectories. This is not merely an academic concern, but a necessary step toward ensuring that these technologies serve the public interest and respect artistic integrity. Rather than passively adapting to a rapidly evolving creative landscape, we must insist on development practices that are inclusive, consent-driven, and open to critical scrutiny. This review aims to contribute to that effort by highlighting key gaps, surfacing under-examined ethical issues, and encouraging more transparent, reflective, and responsible design practices in the field.

9 Conclusion

This review has highlighted the challenges in assessing the responsibility and ethical implications of generative AI systems for music. Whilst many of these models, on the surface, may seem to demonstrate capabilities to the layperson may seem impressive, for those with interest in deeper working with these tools, their opaque nature, proprietary constraints, and varied licensing structures make it difficult to fully dissect their impact on creative agency, transparency, and artistic control. Many AI-driven music tools provide limited insight into their datasets, biases, or decision-making processes, making it challenging for practitioners to critically engage with these technologies beyond surface-level use.

To address these gaps, the authors propose that future work should prioritise the development of clear evaluation frameworks tailored to creative domains, specifically those that focus on bring people back into AI, such as the work discussed by Bryan-Kinns et al. (2025a). These frameworks could include interactive tools, such as transparency dashboards, dataset visualisations, and user interface journeys that expose how models generate content and where limitations or ethical concerns arise. Such tools should be co-designed with artists to ensure they align with the needs of these communities.

Developers and researchers must also advocate for open licensing practices, dataset documentation standards, and greater explainability in model design. By embedding transparency and accountability into the infrastructure of musical AI, we can foster a more equitable and critically engaged relationship between human creativity and machine-generated content.

10 Ethical Statement

Overall, this work aims to adhere to ethical best practices for conducting research in the field of artificial intelligence and musical creativity. In particular, the main ethical considerations for this work are:

- **Use of Publicly Available Information:** The survey is based solely on publicly accessible documentation, academic publications, and official repositories of 27 AI models for music generation.
- **Focus on Responsible AI Principles:** The analysis outlines Responsible AI attributes, including transparency and explainability, fairness, accountability, and ethical use. By highlighting gaps in these areas, the study seeks to support more equitable and reflective development of AI systems in music-making.
- **Critical Engagement with Ethical Gaps:** The research identifies and discusses deficiencies in transparency regarding model training data, licensing, source code availability, and ethical reporting. These issues are presented to call attention to broader structural challenges in AI development for creative practice. Overall, the work aims to encourage the development of clearer evaluation frameworks that support musicians, producers, and developers in understanding the capabilities and limitations of AI systems used in creative contexts.

Acknowledgments and Disclosure of Funding

This project was funded by Responsible Artificial Intelligence (RAI) UK International Partnerships (UKRI EPSRC grant reference EP/Y009800/1)

References

Agostinelli, A., Denk, T. I., Borsos, Z., Engel, J., Verzetti, M., Caillon, A., Huang, Q., Jansen, A., Roberts, A., Tagliasacchi, M., et al. (2023). Musiclm: Generating music from text. *arXiv preprint arXiv:2301.11325*.

AIMI (2024). Aimi. https://www.aimi.fm/ [Accessed: 03-02-2025].

AIVA (2019). Aiva. https://www.aiva.ai/ [Accessed: 05-02-2025].

Amadeus (2019). Meet your artificial intelligence powered songwriting assistant - amadeus code. https://amadeuscode.com/app/en [Accessed: 05-02-2025].

BELIBOU, A. and IFTENE, L. (2024). Artificial intelligence and music. a literature review. *ICT in Muzical Field/Tehnologii Informatice si de Comunicatie in Domeniul Muzical*, 15(1).

BOOMY (2019). Boomy - make generative music with artificial intelligence. https://boomy.com/ [Accessed: 05-02-2025].

Bryan-Kinns, N. (2024). Reflections on Explainable AI for the Arts (XAIxArts). *Interactions*, 31(1):43–47.

Bryan-Kinns, N., Sutskova, O., Fiebrink, R., Perry, P., Wilson, E., Wszeborowska, A., et al. (2025a). Bringing people into ai. University of the Arts London. DOI: https://doi.org/10.58129/9mpb-h282.

Bryan-Kinns, N., Wszeborowska, A., Sutskova, O., Wilson, E., Perry, P., Fiebrink, R., Vigliensoni, G., Lindell, R., Coronel, A., Correia, N. N., et al. (2025b). Leveraging small datasets for ethical and responsible ai music making. In *ACM AudioMostly AM.ICAD 2025*.

Bryan-Kinns, N., Zhang, B., Zhao, S., and Banar, B. (2024). Exploring Variational Auto-encoder Architectures, Configurations, and Datasets for Generative Music Explainable AI. *Machine Intelligence Research*, 21(1):29–45.

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

Chen, J. and Du, R. (2021). Music generation using deep learning with spectrogram analysis. In *2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, pages 589–595. IEEE.

Copet, J., Kreuk, F., Gat, I., Remez, T., Kant, D., Synnaeve, G., Adi, Y., and Défossez, A. (2024). Simple and controllable music generation. *Advances in Neural Information Processing Systems*, 36.

Evans, Z., Parker, J. D., Carr, C., Zukowski, Z., Taylor, J., and Pons, J. (2024). Stable audio open. *arXiv preprint arXiv:2407.14358*.

Feuerriegel, S., Hartmann, J., Janiesch, C., and Zschech, P. (2024). Generative ai. *Business & Information Systems Engineering*, 66(1):111–126.

Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., Falco, G., Fiore, S. M., Garibay, I., Grieman, K., Havens, J. C., Jirotka, M., Kacorri, H., Karwowski, W., Kider, J., Konstan, J., Koon, S., Lopez-Gonzalez, M., Maifeld-Carucci, I., McGregor, S., Salvendy, G., Shneiderman, B., Stephanidis, C., Strobel, C., Ten Holter, C., and Xu, W. (2023). Six Human-Centered Artificial Intelligence Grand Challenges. *International Journal of Human–Computer Interaction*, 39(3):391–437.

Google (2018a). Magenta continue. https://magenta.tensorflow.org/studio/#continue [Accessed: 05-02-2025].

Google (2018b). Magenta drumify. https://magenta.tensorflow.org/studio/#drumify [Accessed: 05-02-2025].

Google (2018c). Magenta generate. https://magenta.tensorflow.org/studio/#generate [Accessed: 05-02-2025].

Hagendorff, T. (2024). Mapping the ethics of generative ai: A comprehensive scoping review. *Minds and Machines*, 34(4):39.

Harmonai-org (2022). Dance diffusion. https://github.com/Harmonai-org/sample-generator [Accessed: 03-02-2025].

Herremans, D., Chuan, C.-H., and Chew, E. (2018). A Functional Taxonomy of Music Generation Systems. *ACM Computing Surveys*, 50(5):1–30.

Ho, J., Chan, W., Saharia, C., Whang, J., Gao, R., Gritsenko, A., Kingma, D. P., Poole, B., Norouzi, M., Fleet, D. J., et al. (2022a). Imagen video: High definition video generation with diffusion models. *arXiv preprint arXiv:2210.02303*.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851.

Ho, J., Saharia, C., Chan, W., Fleet, D. J., Norouzi, M., and Salimans, T. (2022b). Cascaded diffusion models for high fidelity image generation. *The Journal of Machine Learning Research*, 23(1):2249–2281.

Huang, Q., Park, D. S., Wang, T., Denk, T. I., Ly, A., Chen, N., Zhang, Z., Zhang, Z., Yu, J., Frank, C., et al. (2023). Noise2music: Text-conditioned music generation with diffusion models. *arXiv preprint arXiv:2302.03917*.

Kashyap, R. and Kumar, A. (2019). *Challenges and applications for implementing machine learning in computer vision*. IGI Global.

Liu, H., Huang, R., Liu, Y., Cao, H., Wang, J., Cheng, X., Zheng, S., and Zhao, Z. (2024). Audiolcm: Text-to-audio generation with latent consistency models.

Loudly (2023). Ai music for your creative universe | loudly. https://www.loudly.com [Accessed: 05-02-2025].

McCallum, L. (2019). Magnet (github repository). https://github.com/Louismac/MAGNet/ [Accessed:06-02-2025].

McCallum, L. (2023). Dorothy (github repository). https://github.com/Louismac/dorothy [Accessed:07-04-2025].

McPherson, A. P., Jack, R. H., Moro, G., et al. (2016). Action-sound latency: Are our tools fast enough? In *NIME 2016*.

Mehri, S., Kumar, K., Gulrajani, I., Kumar, R., Jain, S., Sotelo, J., Courville, A., and Bengio, Y. (2016). Samplernn: An unconditional end-to-end neural audio generation model. *arXiv preprint arXiv:1612.07837*.

Morreale, F. (2021). Where does the buck stop? ethical and political issues with ai in music creation. *Transactions of the International Society for Music Information Retrieval*, 4(1):105–113.

Morreale, F., Bahmanteymouri, E., Burmester, B., Chen, A., and Thorp, M. (2024). The unwitting labourer: extracting humanness in ai training. *AI & SOCIETY*, 39(5):2389–2399.

Mubert (2016). Mubert ai music generator — royalty free music. https://mubert.com/ [Accessed:03-02-2025].

Nayar, V. (2025). The ethics of ai generated music:: A case study on suno ai. *GRACE: Global Review of AI Community Ethics*, 3(1).

Newman, M., Morris, L., and Lee, J. H. (2023). Human-ai music creation: Understanding the perceptions and experiences of music creators for ethical and productive collaboration. In *ISMIR*, pages 80–88.

Oppenlaender, J. (2022). The Creativity of Text-to-Image Generation. In *Proceedings of the 25th International Academic Mindtrek Conference*, Academic Mindtrek '22, pages 192–202, New York, NY, USA. Association for Computing Machinery.

Piskopani, A. M., Chamberlain, A., and Ten Holter, C. (2023). Responsible AI and the Arts: The Ethical and Legal Implications of AI in the Arts and Creative Industries. In *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*, pages 1–5, Edinburgh United Kingdom. ACM.

Riffusion (2019). Riffusion – about. web.archive.org/web/20221216092717/https://www.riffusion.com/about [Accessed:03-02-2025].

Sarkar, A. (2023). Exploring Perspectives on the Impact of Artificial Intelligence on the Creativity of Knowledge Work: Beyond Mechanised Plagiarism and Stochastic Parrots. In *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work*, CHIWORK '23, pages 1–17, New York, NY, USA. Association for Computing Machinery.

Song, Y. and Ermon, S. (2019). Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32.

Soundful (2023). #1 ai music studio - ai music generator for creators | soundful. https://soundful.com/ [Accessed: 05-02-2025].

Soundraw (2024). Ai music generator soundraw. https://soundraw.io/ [Accessed: 05-02-2025].

Splash (2023). Splash - imagine music. https://www.splashmusic.com/ [Accessed: 03-02-2025].

Suno (2023). Suno. https://suno.com/ [Accessed: 03-02-2025].

Udio (2024). Udio ai music generator - make original tracks in second. https://www.udio.com/home [Accessed:03-02-2025].

Vigliensoni, G., McCallum, L., Maestre, E., Fiebrink, R., et al. (2022). R-vae: Live latent space drum rhythm generation from minimal-size datasets. *Journal of Creative Music Systems*, 1(1).

Wilson, E., Schubert, D., Satomi, M., McLean, A., Amaya Gonzalez, J., et al. (2023). Mosaick: Staging contemporary ai performance-connecting live coding, e-textiles and movement. In *Proceedings of the 7th International Conference on Live Coding (ICLC 2023)*.

Xambó, A. (2021). Virtual agents in live coding: A short review. *arXiv preprint arXiv:2106.14835*.

Yang, D., Yu, J., Wang, H., Wang, W., Weng, C., Zou, Y., and Yu, D. (2023). Diffsound: Discrete diffusion model for text-to-sound generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.