

**AIMC 2023**

# **Silicon for Orchestra and Artificial Intelligence**

**Robert Laidlow<sup>1</sup>**

<sup>1</sup>Jesus College, University of Oxford

**Published on:** Aug 29, 2023

**URL:** <https://aimc2023.pubpub.org/pub/wbt654g2>

**License:** [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

## Silicon for Orchestra and Artificial Intelligence

### **ABSTRACT:**

This paper discusses the recent large-scale orchestral work ‘Silicon’, composed for the BBC Philharmonic. ‘Silicon’ uses and examines artificial intelligence in the context of orchestral music in three key ways, divided between its three movements: as a symbolic-generative ‘composer-like’ agent, as a tone-transfer ‘instrument-like’ agent, and as a neural synthesis ‘performer-like’ agent. These strategies are discussed alongside novel compositional ideas discovered through working with AI, and the broader question of how a performing institution, such as an orchestra, might incorporate advanced technologies into its fabric. The paper situates AI as a useful and creative tool for composing new music for orchestra, and the orchestra as a useful and creative tool for examining AI.

**KEYWORDS:** Orchestra, Performance, Creative Process, Neural Synthesis, Symbolic AI, Tone-Transfer

## **1 Introduction**

### **1.1 Overview**

This paper is an explanation and illumination of one artistic process that utilises artificial intelligence (AI) to compose the large-scale orchestral work ‘Silicon’. It focusses on the artistic possibilities afforded by three distinct categories of AI, which I have termed ‘composer-like’, ‘instrument-like’, and ‘performer-like’.

These distinctions of musical AI are utilised respectively in ‘Silicon’'s three movements, titled ‘Mind’, ‘Body’, and ‘Soul’. An overview of each algorithm used is given, in addition to a description of a musical strategy employed to utilise each effectively within a performing orchestra context.

The aesthetic questions that these types of AI produce are given equal weight to the technologies themselves, and some possible responses to these questions are explored throughout ‘Silicon’. These questions touch upon the relationship between the past and the future in both AI and music research, on authenticity and ‘fakeness’, and on the essence of an orchestral performance.

At this point it would be appropriate to include a short review of other works for orchestra and AI that ‘Silicon’ builds upon. This is, however, a nascent field with many hurdles to composers – not least access to technology and the support of a major orchestra. Setting aside curiosity projects that use AI to “complete” unfinished music by deceased European composers, there are very few direct comparisons. George E. Lewis provides one recent example with his 2021 work ‘Minds in Flux’ for orchestra and interactive electronics, in addition to his earlier ‘Virtual Concerto’. Smaller scale works utilizing AI are much more common, with recent examples including ‘This is Fine’ (Salem, 2021), ‘That’s what they said’ (Ma, 2022), and ‘Love Letters [With AI]’ (Oliver, 2023).

‘Silicon’ does, however, join a much larger collection of works exploring the orchestra and electronics more generally. This area is too diverse and established to provide a complete overview here, but some examples relevant to the compositional process of ‘Silicon’ include the work of Saariaho (e.g., ‘du cristal’ 1989), Steen-Andersen (e.g., ‘Double Up’, 2014), and Luther Adams (e.g., ‘Dark Waves’, 2007).

Visit the web version of this article to view interactive content.

Robert Laidlow (2022): Silicon

## 1.2 Defining Composer-, Performer- and Instrument-like AI

My use of the terms composer-like, performer-like, and instrument-like AI builds upon an existing distinction between “performance-time” and “design-time” algorithms (Fiebrink & Caramiaux 2018). These two existing terms differentiate algorithms that are intended to be used during a live performance, and those that help a composer design a performance.

In this paper I use a further distinction to the idea of “performance-time”, highlighting the difference between an instrument-like AI and a performer-like AI in a live music-making setting. In the case of the compositional process for ‘Silicon’, this was an important distinction. Since the music was to be performed live, there is a material difference between a human interpreting written music *using* an instrument, and an AI algorithm acting as an independent performer within the fabric of the orchestra.

As I define it, an instrument-like AI operates live in performance, but not autonomously. They are used by a human performer to give a performance (in the case of ‘Silicon’, to interpret notated music). While instrument-like AI will often have interactive, adaptive, or responsive elements using AI technology, salient musical parameters (such as pitch, rhythm, dynamic, or timbre) will be controlled by a musician. The example in this paper is DDSP-VST (Engel et al 2020), released in 2022. Other recent examples might include the work of the Intelligent Instruments Lab, RAVE (Caillon & Esling 2021) or the Wekinator (Fiebrink & Cook 2010). Like traditional acoustic instruments, then, instrument-like AI might be considered extensions of a human performer.

A performer-like AI, on the other hand, operates on its own terms, in a way that can feel improvisational. They are not directly controlled by a human player, but instead make their own autonomous decisions which may or may not be informed by the actions of other (human) performers alongside them. A good example is the work of George E. Lewis on his Voyager system (Lewis 2000). Other examples of artists creating or collaborating on their own performer-like AI include Holly Herndon in her 2019 album PROTO, Jennifer Walshe and Memo Atken in their 2018 work Ultrachunk, and Dadabots, which was originally based on SampleRNN (Mehri et al 2017). Typically, performer-like AI have generated audio, though symbolic-generative AI should not be excluded (perhaps they could then use an instrument-like AI to render their generations into sound).

Perhaps the key differences between instrument-like AI and performer-like AI are in the direction of attention during a performance. Instrument-like AI will tend to pay attention only to the player directly controlling it, receiving instruction only from them; performer-like AI will pay attention to the group at large, or perhaps to nobody at all. When an instrument-like AI is employed, other performers will pay attention to the human controlling it for musical cues; when a performer-like AI is used, other performers will listen directly to that algorithm (if appropriate for the given musical style), treating it as a group member.

With these two distinctions already made, I have found it useful to reterm the existing concept of “design-time” algorithms as composer-like. In the context of ‘Silicon’, composer-like AI informs the underlying musical material of the work either directly (through generating music that is to be performed by musicians) or indirectly (through challenging the composer to think in new ways). If the creative process is imagined as a series of links in a chain, each of which is a creative act, then these composer-like AIs can fill a similar role in that chain to humans. ‘Creative links’ might be things like coming up with initial inspiration for the work, basic musical ideas, developments of earlier ideas, structural plans, or descriptions of soundworlds. While instrument-like and performer-like AI will be apparent on the surface of a performance – they will be literally audible – composer-like AI will often be less obvious.

This view draws upon recent research into distributed creativity: Juliet Fraser’s (2019) recasting of composer and performer to ‘agent[s] in the process of creating the work’ and Jennifer Torrence’s (2018) proposed a model of creativity that treats performers as a ‘deviser’ where ‘both parties contribute to creative and practical decision making’ are both foundational ideas. Cassandra Miller (2018) and Zubin Kanga (2014) also both describe collaborative processes that involve the sending of material from one collaborator to another, and the subsequent analysis or editing of that material according to each person’s expertise or subjective sensibilities. Examples of composer-like AI might be MusicLM (Agostineli et al 2023), Jukebox (Dhariwal et al 2020), or MuseNet (Payne 2019), the latter of which is discussed in this paper.

Many algorithms might fit into multiple of these categories, or none at all, depending on who is using them at what point in the creative process they are being used, and what is being done to whatever the AI generates. This paper is not intended to provide an inflexible framework for imagining how AI fits into a creative framework, but rather a starting point for discussion based on recent artistic work undertaken. In my own experience, I have found this categorisation more helpful than both the design/performance dichotomy and the more technical audio/symbolic-generative dichotomy.

This overall approach to utilising algorithms is what has been usefully described as “human-centred” (Fiebrink & Caramiaux 2018) – algorithms that support the creative process, rather than a different type of creative process which focusses on creating rules for the algorithms that will generate sound. Since machine learning is often unsupervised, especially the tools I have used, there is relatively little scope for telling the algorithm what to do in any case – machine learning forms the crux of Fiebrink & Carmaiaux’s argument for human-centrism, perhaps for this reason.

I have found that a human-centred view transforms algorithms from tools to solve problems into imperfect mirrors that can “help users express hidden, ill-formulated ideas” (Pachet 2008). While Pachet is not referring to machine learning here, this possibility seems to me even more important in machine learning because it learns for itself how to form this mirror, thus revealing elements the user may not have previously considered.

A human-centred approach to AI algorithms also encourages imperfection, since I can consider the results of algorithms (or even the idea of an algorithm) to be compositional material, rather than completed music. Research that prioritises a human-centred approach often references the possibilities of using ‘bad’ algorithms with artefacts, aberrations, or other ‘unwanted’ results (Wiggins 2018). Such an approach has also recently underpinned some composers’ work in highlighting unintended bias of machine learning algorithms in wider society (Criado Pérez 2020; Dastin 2022; also discussed in Ma 2021).

To summarise, I propose that the existing term ‘performance-time’ which is applied to algorithms might be usefully divided into instrument-like and performer-like, depending on what role the composer is using that algorithm to fulfil. This has been, to me, more useful in imagining how to embed AI technology within an artistic practice. The division into instrument-like or performer-like is dependent on the compositional process, and not necessarily the algorithm itself. A composer might use the same algorithm as instrument-like in one piece, and performer-like in another, or even within the same work. To fit the same framework, I have also redefined the existing term ‘design-time’ as composer-like. Composer-like AI inform underlying musical materials, but may not always be immediately audible in the performance itself.

### 1.3 Algorithmic Time

Some of the music in this paper is inspired by ideas of ‘algorithmic time’. Rohrerhuber (2018) asserts that “algorithmic methods suggest a break with the idea of time as an immediate grounding” because actual time (the kind that is measured on a clock) is a less effective measure of progress than observing which step an algorithm has reached in its process. An algorithm therefore contains its own time – which it procedurally unfolds step-by-step – that does not have a direct relationship with actual time. Grounding a musical work in algorithmic time, while the human listener or performer necessarily exists in actual time, encourages investigation into scale and linearity, two areas of interest in for this paper and in the research of others (Spiegel 1981; Magnussen & McLean 2018).

Rohrerhuber goes on to state that “eventually algorithmic music will turn out to be not only affected by how we understand temporality, but also it will turn out to be a possible method to constitute and convey the peculiar existence of time”. The movements of ‘Silicon’ accordingly approach musical time as a dimension that can be manipulated, expanded, contracted, or otherwise developed.

## 2 ‘Mind’, Time, and Composer-Like AI

## 2.1 MuseNet and Music of the Past

To explore AI acting in a composer-like capacity (that is, used to inform underlying musical material) in ‘Mind’, I employed MuseNet (Payne 2019). MuseNet is a general-purpose AI that employs a transformer architecture to generate symbolic (MIDI) musical data. During training, each file (also MIDI) in the dataset is tagged with its composer or genre. This allowed MuseNet to learn the musical fingerprints, or at least what it deemed to be the musical fingerprints, of many different composers. The user can specify which composer or genre MuseNet should emulate when generating MIDI.

For this movement, I instructed MuseNet to generate in the style of Mozart. This suited the overall theme, discussed presently, of creating something new out of something old and allowed me to showcase the algorithm when performing at its intended task (stylistic composition) as well as when pushed into new places.

Like many contemporary algorithms, MuseNet uses existing music both as a dataset from which to learn musical rules and as a yardstick against which to judge the quality of generations. Marcus du Sautoy (2020) writes that ‘Bach is the composer most composers begin [learning] with, but he is the composer most computers begin with too’, and indeed Bach is often the choice of dataset and generation for much recent research into composer-like AI (e.g., Hadjeres et al 2017; Fang et al. 2020; Whorley & Laney 2021). Occasionally this has led to the impression that one of the main uses of composer-like AI might be to complete unfinished pieces by dead composers, as indeed we have seen in recent years with AI ‘completions’ of Beethoven and Schubert (see Goodyer 2021).

Classical music, and by extension the orchestra as an institution, is also well-known for using the past (specifically its own past – European art music) to create the present and the future. The performance of established Western music is perhaps the genre’s defining trait, which is evident in the programming of symphony orchestras (Donne, *Women in Music* 2020; Gotham 2014). Why do we do this? My most optimistic view is that it’s because we believe that ideas from the past can have something to say in the present – something beyond merely being a benchmark by which to judge technical progress. Conversely, modern composers often use references to older music, or different genres of music, to make exciting and fascinating musical arguments.

The piece, therefore, is intended to examine questions of legitimacy in addition to challenging how the compositional process is affected through using a composer-like AI. How does the past legitimise the future in some AI research and in some orchestral music, and how can accepting or disrupting this legitimisation form the basis of a new piece of music? It is important to note that the music of the past that both orchestras and AI research has tended to engage in is Western classical music, which cannot be described as a stand-in for global musical culture.

## 2.2 Form and Interlocking

I chose to compose ‘Mind’ using sonata form as a departure point (Figure 1), which I hypothesized would be well known both to the algorithm (having been trained on many sonatas) and some listeners with specialized knowledge. I was interested in exploring and showcasing the change from familiar to uncanny, and in slowly polarising the AI-generated material until it ended up in a very unfamiliar sonic landscape. I felt the historic sonata form was a useful vessel through which to explore advanced technology’s relationship with music of the past.

Figure 1: Structural diagram of ‘Silicon Mind’. Top shows sonata form relation. Section containing interlocking show with bracket. Red arrows denote locations of axes of reversal. Bottom shows where exposition is heard forwards and, later, reversed.

I employed what I term ‘interlocking’ as my primary strategy for working with MuseNet generations within this form. Interlocking refers to the practice of alternating blocks of AI- and human-composed musical material. At the time of composing ‘Mind’, machine learning could not produce coherent music of longer than around 30 seconds (Dhariwal et al 2020). I originally developed interlocking to mitigate this structural problem, by using my own material to re-orient the music, but it quickly became useful in many other ways.

The ‘exposition’ of the piece is created using this interlocking technique (Figure 1). I then developed the ideas generated through interlocking in various compositional ways throughout the rest of the work. This included being accelerated, slowed down, microtonally transposed, reversed, rotated harmonically and elongated at different moments in the piece.

## 2.3 Musical Time

Somewhat more speculatively, I also wanted to investigate what music of the future might sound like if an AI wrote music that was not bound to rules of the past. According to the definition of algorithmic time given in the introduction, time ‘begins’ with the first step and ‘ends’ when the algorithm is complete. It does not necessarily matter how long, in actual time, these steps take. This is an idea that chimes with some orchestral music. I imagined that a sonata form could be considered a type of algorithmic time, where it is at least as informative to understand the relationship of the internal sections that unfold in a specific order than it is to count how many seconds have passed in actual time.

This kind of step-based time exists in any algorithm, but ‘Mind’ also has major elements of musical time manipulation derived specifically from AI algorithms. AI algorithms learn from audio data like WAV files, or from symbolic data like MIDI files. Whether audio or MIDI, for AI training purposes this data can be transformed into an image, such as a spectrogram or a MIDI roll (Carykh 2017).

Images do not exist in time – they are static. It’s only when we tell something to play that image from left to right that the dimension of time suddenly originates. But if a machine was creating music for itself, in a

hypothetical future where machines exist that enjoy listening to music for its own sake, musical time probably wouldn't need to work in the way we hear it. The image-music could be enjoyed all at once, top-to-bottom, right-to-left or the traditional start-to-end. I wondered if this might be an equally valid way of hearing AI-generated music, even if it made less intuitive sense to a listener.

To enact this in 'Mind' I created several axes of reflection across the piece (Figure 1). On either side of these axes, the same music is heard both forwards and backwards. This is not only retrograding rhythms and pitches, but also the timbre, decay, and attack of the sound. I imagined reading a spectrograph backwards (right-to-left) and therefore reversing the entire sound. Often the entire orchestra is not in reverse, but rather some instruments flip at an axis of reflection, while others continue moving through the warped sonata form.

Realising this with only the physical instruments of the orchestra presented an enjoyable challenge. Many instruments can emulate a near-enough reversed sound by simply starting quietly and cutting off any resonance at the end of the note. Some required more thought; reversing the sound of the vibraphone, for example, requires the percussionist to first bow the note to produce a sustained note, before striking and damping the note with a mallet. For certain instruments such as gongs (which produce a very different timbre when bowed to struck), I emulated decay through other instruments in the orchestra (Figure 2).

Figure 2: Example of emulating decay for percussion instruments. In this case the bass clarinet is used to provide resonance for tuned gongs

The final way I wanted to explore a kind of AI-algorithmic form was through an idea of branching narratives. MuseNet can be instructed to create any number of responses to a prompt, which will all be created simultaneously, and each will be different. Several times while composing, I orchestrated and included one MuseNet generation before rewinding back to the start of that generation to use another, creating a sonification of constant progress through many iterations of the same task. This allowed me to either show two wildly different responses to the same musical idea or orchestrate two responses to the same prompt in different ways (Figure 3).

Figure 3: Two different orchestrations of material separated by ratchet (Bar 35) which reverts the music to the beginning of the phrase

### **3 'Body', Authenticity, and Instrument-Like AI**

#### **3.1 Fake and Real**

The second movement, 'Body', is concerned with ideas of authenticity. In recent years we have become familiar with AI's capacity for creating believable fakes. This technology is used to automatically generate stories that resemble human-written news and by social media giants to encourage engagement, with the dissemination and promotion of fake news stories as known by-product (Wang et al 2018). It is now a regular



occurrence to see AI algorithms used to create fake videos showing public figures in unfavourable light (Botha & Pieterse 2020) and it has also been used in movies to allow deceased actors to appear in new releases (e.g. Peter Cushing and Carrie Fisher in *Rogue One: A Star Wars Story*) or to de-age live ones (Sargeant 2017).

Accordingly, we are now becoming used to questioning the provenance of believable-looking sources. Real is not, however, necessarily the same as authentic. Authenticity is a much more subjective question – one with which classical musicians are familiar. Discussions and disagreements emerging around, for example, performing Bach on the modern piano (Edidin 1998) or using vibrato during 18<sup>th</sup> century symphonies (Norrington 2004) can be viewed, at least partially, as questions of authenticity.

This leads to questions explored in ‘Body’. What exactly is fake music? And does fake or inauthentic music become any more authentic when performed by an orchestra – by real people? Perhaps most importantly, I wanted to hear what this ‘fake sound’ technology sounds like. I wanted to embed an instrument-like AI – DDSP-VST - within the orchestra, to be played by an orchestral musician, as a model for how orchestras might be constituted in the age of AI.

One research paper that particularly interested me showcasing deepfake technology was ‘Everybody Dance Now’ (Chan et al 2019). It demonstrates taking a video of a dancer (Source), an image of a second person (Target), and the use of AI to make the Target appear to move like the Source. To do this, it strips the Source video down to a basic set of moving points and lines, abstractly representing the human body. With this distilled from the Source, the AI then rebuilds the video, this time with the Target fleshing out the skeletal nodes. I found the way that computer vision ‘sees’ people fundamentally differently to how we see people fascinating – and a perhaps a little unnerving. An answer to what fake music might sound like lay, for me, in the relationship between the surface – the Target – and the hidden layers – the Source.

### 3.2 Sources

*Silicon Body* has a Source, a layer of music that sits underneath and drives the whole piece. This skeletal musical framework is made up of three simple alternating patterns of pitches and rhythms. The Source layer moves through and mutates these ideas in turn, with the aim of realising two of them at the same time. At this point, its logical argument finishes, and the movement finishes soon after.

The Source is performed on an instrument-like AI: Google Magenta’s DDSP-VST (Engel et al 2020). DDSP-VST can transform one instrument’s (an audio source) sonic content into another (an audio target) in real time through training deep learning models on target audio data. I worked with Magenta to create models based on the recordings of my performer colleagues (across a range of orchestral wind and string instruments) that DDSP-VST could learn from and imitate during ‘Body’. Each recording was around 30 minutes long, consisting of free improvisations and musical extracts chosen by the performer, and each resulted in one DDSP-VST model.

DDSP-VST was then employed, compositionally, in two key ways. Towards the start of the movement, DDSP-VST would often mimic the exact instrument playing in the music at the time. For example, if a clarinet was playing, I would use the clarinet model at that moment. DDSP-VST slips in and out of audibility, sometimes relying on the physical theatre of *hearing* a clarinet, but seeing that no player is performing at that time. Later in the piece, I made heavier use of DDSP-VST's ability to differentiate between “harmonic” and “noise” content, which it understands to be extraneous sounds in its training data such as violin bow noises and breathing. Harmonic and melodic content was transformed into purely “noise” content, and several models were also layered on top of each other to create “noisy” textures. In this performance, models were changed automatically in a predetermined pattern depending on which key the player was pressing. The software was relatively volatile at the time of composition, so this decision was made to reduce the chance of computer failure mid-performance. In future performances, it might be interesting to randomise which model(s) are being used at any given moment.

The orchestral keyboardist performs on a digital piano, which is linked to DDSP-VST through Ableton Live (Figure 4). The sound produced is amplified through a speaker close to the keyboardist, localizing the sound to that player. I found DDSP-VST uniquely useful among instrument-like AI, because of how simple it is to provide to a musician who might have little to no experience of ever using a computer in performance. In this music, I was keen to give orchestral performers complete control over AI tools where possible but was mindful of needing to provide them with interfaces (such as pianos and sheet music) that they would be most familiar with.

Figure 4: DDSP-VST Interface on Ableton Live. Instrument model is selected at the top. The middle of the figure shows the preferred range and dynamic of the model. The further outside of the highlighted box the input note is, the more the model is required to “guess” what that note should sound like when realized using the target model. The bottom row shows adjustable parameters, including the proportion of “harmonic” and “noise” content as discussed above.

### 3.3 Targets

On top of this Source are superimposed three Target styles of music that are performed by the orchestral instruments. Inspired by ‘Everybody Dance Now’, these three styles are based on different types of dance music – big band jazz, electronic techno, and folk. I composed the jazz and techno styles in their entirety, in short score (Figures 5 and 6).

Figure 5: Jazz Style Short Score

Figure 6: Grid of Musical Material used in Techno Style

After composing and discarding a short score for the folk-dance style, I decided to utilise Folk-RNN for its material (FolkRNN.org). It seemed to be fitting to use AI to generate some of the music. After generating a

tune that I deemed would fit with the wider piece, I wanted to push it into uncanny territory. I did this through, for example, applying microtonal glissandi to the melody (Figure 7).

Figure 7: Microtonal Manipulation of Folk-RNN Melody

*Silicon Body* essentially consists of each of these three Target dance styles occurring simultaneously, but only one is usually heard at any given moment. The surface of the music cuts between the three styles at an ever-increasing rate, rotating through them faster and faster until the music reaches a perceived breaking point. At this point, the same point that the Source concludes its logical argument, the Source is revealed on its own and the piece then ends. Each style has their own set of orchestral instruments that do not usually overlap – the only instrument they share is DDSP-VST.

The Source layer determines most musical properties of the Target layer, including tempo, tonal area, and rhythm, forcing the Target styles to be transposed, stretched, crushed, or become otherwise ‘fake’ according to the Source layer logic (Figure 8).

Figure 8: Orchestral music (Target) parameters such as tonal centre, dynamic, and tempo are dictated by DDSP-VST (Source)

## 4 ‘Soul’, Secrets, and Performer-Like AI

### 4.1 PRiSM-SampleRNN

In the third movement, ‘Soul’, I intended to experiment with the PRiSM reimpementation of SampleRNN, ‘PRiSM-SampleRNN’ (Melen 2020), as a performer-like AI – that is, as a sounding part of the orchestra on stage, rather than part of the compositional process or controlled by a human performer. This being an early major test of this algorithm, PRiSM-SampleRNN's developer Christopher Melen and I wanted to challenge its ability to deal with larger dataset size and higher audio quality.

Until this point, most testing of PRiSM-SampleRNN had utilised a training dataset of approximately 0.5 to 10 hours in length. The BBC Philharmonic provided me with access to their archive of broadcasted concerts, which I turned into a dataset lasting approximate 2000 hours. Additionally, PRiSM-SampleRNN had mostly been used to generate relatively low-quality audio, at 16,000Hz. This is because a lower sample rate provides a host of benefits, primarily making testing, training, and generating audio much faster. The model for *Silicon Soul* was trained to produce audio at 44,100Hz (CD quality) instead.

When training (see Appendix for technical details) was finished, we had five different models, representing different stages of the training process. Each model had its own ‘sound’ – its own take on how to imitate the BBC Philharmonic (Figure 9). In my own opinion, the boost in audio quality not only resulted in higher fidelity audio generations, but also a higher musical quality; the generations seemed more assured, contrapuntal, and consistent than lower fidelity testing.

AI Model Name	Notes
BBC-Full Epoch 1	<p>Very subtle at lower temperatures, possibly useful as a morphing background texture. Timbrally sometimes half sine-wave, half orchestral sounds.</p> <p>Recognisable progression towards very symphonic (Straussian) brass sounds as temperature is increased. Loud strings, little woodwind.</p>
BBC-Full Epoch 2	Fuzzy and volatile. Very low audio quality, with booming and clipping bass. Higher temperatures generally slightly higher quality. Unlikely to use Epoch 2 generations in this piece.
BBC-Full Epoch 3	Lower temperature (0.9-0.95) has beautiful and calm textures, quite ethereal. 0.975 temperature retains this quality but with occasional flashes of recognisable orchestral activity. Some generations transform into or out of applause.
BBC-Full Epoch 4	Very slow moving and languid. Mysterious in places – good for supporting orchestra or collaging on top of itself. Very high quality audio.
BBC-Full Epoch 5	Very high quality audio again. Epoch 5 generations are in motion, exciting and engaging. At lower temperatures the bass is fuzzy but this goes away from 0.975 onwards. 0.99 temperature generations are very exciting and symphonic – they could be actual recordings. Use for climax of electronics part.
Presenter-Only	<p>Lots of useful material here. Plenty of presenters speaking, they speak in what sounds like a garbled made-up language. Lots of applause and tuning notes, sometimes blending with presenter's voice.</p> <p>After some 40 generations from this dataset, I have never heard the AI imitate a female presenter voice. Interesting comment on the overall trends of BBC Radio 3 archive recordings.</p>

Figure 9: My original notes on each Sample-RNN model (Full orchestral dataset and presenter-only)

My dataset was not only music, but entire broadcasts from BBC Radio 3. This is because the BBC Philharmonic is a broadcast orchestra, whose remit is to provide content for radio. In practice, this meant that my dataset had other radio-like material within it, such as audiences applauding, presenters introducing music, and the orchestra tuning in the background. I realised that if I wanted to make an AI respond specifically to the BBC Philharmonic, I should also include these sounds – they are, in my opinion, part of the DNA of a radio broadcast orchestra. I created a separate dataset of just these non-musical sounds and trained a separate PRiSM-SampleRNN model on it (Figure 10).

The audio generations of these six AI models were used to create a 4-channel accompaniment to the orchestra, one which moves from the non-musical audio results through to volatile and dynamic imitation of full orchestra. The 4 monitors playing this accompaniment surrounded the orchestra, allowing more granular control over the physical locations of the sounds.

PRiSM-SampleRNN generations are monophonic, and if the same settings are used for two generations, they will create two very similar pieces of audio (but never identical). I often collaged two or more similar generations from the same model simultaneously, each in a different channel, to create a shifting, stereo-like effect (Figure 10).

Figure 10: Example of Collaging Sample-RNN Material to Create Electronics for Silicon Soul

[Click here to listen to the stereo bounces of these AI-generated electronics.](#)

## 4.2 Musical Secrets

A second area I wanted to explore with ‘Soul’ was aesthetic. To do this, I imagined a ‘perfect’ audio-generative algorithm as a thought-experiment. The thought-experiment algorithm has no artefacts, and it can achieve whatever musical task we set it. It can analyse any amount of data, unrestricted by hardware limitations, and can produce new data (i.e., music) trivially quickly. It can produce sound indistinguishable from human musicians in any genre, historical period, or ensemble. It can even produce entirely new music by combining existing music in novel ways or identifying gaps in its dataset that have never been exploited. But is it music, or does it only *sound* like music?

Would people accept this music, or do we require some kind of secret ingredient in order to feel a genuine connection with art? We don’t know the overall answer to this question, or even if it can be answered, in this specific instance because AI has not reached the fluency of the thought-experiment, but it’s reasonable to imagine it will. And if I take the view that there is more to music than computer data can communicate, what is that secret? Does it exist inherently within the music, or can this secret be imagined or imposed by the audience? Will AI research, in its pursuit of a systematic, mathematical, and function-based understanding of the world, help us understand what the secret of music is?

This and related questions are already under active consideration from a wide range of artists and academics who have influenced my thought. Federico Campagna argues that embracing a worldview he terms “magic”, informed by elements of spiritualism, mysticism, and religion, can help alleviate the difficulties, both personal and social, inherent in a worldview reliant on data (Campagna 2018). Similarly, the authors of the *Atlas of Anomalous AI* explicitly state their aim to ‘re-mythologise AI in a way that reveals the roots of the technological project in spiritual practices, institutions and frameworks’ (Vickers & Allado-McDowell 2021). George E. Lewis describes a view of improvisation as ‘something essential, fundamental to the human spirit’, before going on to assert that attempting to teach computers to improvise ‘can teach us how to live in a world marked by agency, indeterminacy, analysis of conditions, and the apparent ineffability of choice’ (Lewis 2018).

I set out to provide one response to this question by examining it through the lens of orchestral music. I wondered why audiences still go to see the orchestra today. As the pandemic has shown, it is perfectly possible to livestream performances to tune into from home, and there are also sample libraries that allow us to emulate the orchestral sound without needing any humans at all. What is its secret that compels people to physically come and watch humans make these sounds live?

For me personally, it is in understanding an orchestral performance not primarily as an act of creating sound, but rather as an act of community shared between musicians. I wanted to experiment with including performer-like AI inside such a framework.

One strategy of doing this was to make the AI personal to *that* orchestra, as described above through specific use of dataset. In this instance AI is used as a tool to increase the personalisation and site-specific nature of a piece, rather than as a tool to make general rules about music. It is in service of defining what the nature of *this* ensemble is, and in using it I was challenged to make decisions about how to treat the similarities and differences in sound between the physical orchestra and its AI doppelganger.

I also set out to make the orchestral parts ‘site-specific’, similar to how the training data was ‘site-specific’. I wanted each note to make sense only on the instrument it was written, and in that way for the material to be tied to the essence of each sound. Each instrument was given a note that seemed ‘natural’ to that instrument, such as open strings, natural harmonics, notes that sat firmly within the instrument’s ideal playing range, or those with special significance (such as the oboe’s A natural ‘tuning’ note) (Figure 11).

Figure 11: Notes assigned to orchestral instruments. Any brass harmonics are natural (no valves). String notes are either open strings or natural harmonics

Expanding this idea of ‘natural’ sounds, I searched for non-pitched sounds that could only be made by specific instruments in the orchestra. Working with some performer colleagues, I found a series of non-pitched sounds in the strings, percussion, and brass with which to begin the piece that accompanied the fuzzy quality of the PRiSM-SampleRNN towards the start of the electronics track. The orchestra and the AI both act like ‘shells’ of

a regular orchestral performance at various points in the piece: the orchestra through retaining only extended performance techniques, and the AI through acting as a strange audible mirror.

For the majority of the piece, each instrument only plays its assigned note or technique, except in instances where the orchestra specifically imitates material generated by PRiSM-SampleRNN.

## 5 Brief Concluding Thoughts

Through this description of ‘Silicon’, I hope to demonstrate some creative possibilities I have found in AI to both other composers and those who develop algorithms. The distinction into composer-like, instrument-like, and performer-like AI was extremely useful in the development of this work as a means of understanding the role each AI played within a wider creative process. While there is no guarantee that such a framework is generalizable, it may provide a starting point for composers to work with, or to constructively reject.

This distinction also raised many questions which could be areas for future work. The three groups of instrument-like, performer-like and composer-like AI intersect in many places. A composer might use composer-like AI to inform music performed on instrument-like AI, or vice versa. Does this make the instrument-like AI also a composer-like AI, in the end? What about algorithms that generate, for example, notated music live which is to be sight-read by a human performer? The blurry boundaries at the edges of these definitions represent to me some of the most exciting areas of this research as it stands. While these definitions may help with understanding some AI’s role in a creative process, they should not come at the expense of imagining this technology as filling a role wholly new.

There is room for more work at the specific interaction of AI and the orchestra. First and foremost, it would be a great pleasure to see any of the research or views in this paper challenged by other artists utilising AI in their creative process when writing for orchestra. There are still very few composers working in this field, likely due to significant collaboration between cultural and research institutions being necessary to realise this work. It seems that there is an urgent need for further artistic research as a means of understanding what place this technology might take in the wider cultural ecosystem.

Some of this work responds to ideas of music of the past. In the case of ‘Silicon’, that past music has been Western European orchestral music, and this is also true of many other music-AI projects. This music is not representative of all music, and more work needs to be done in opening the field to other practices. Working with non-European music and musicians will not only inform research into whether AI can be generally useful (or not) across a wide range styles, but may also highlight implicit biases towards Western music embedded within AI architecture itself (for example, a focus on equal temperament and metered time). This might, in turn, lead to the development of composer-like, instrument-like and performer-like AI algorithms built upon different fundamental principles.

Throughout the composition and rehearsal process, it is clear that large-scale composer-like AI systems dealing specifically in symbolic or notated material are lagging several years behind those which generate raw audio. By allowing this area to stagnate, researchers are potentially ignoring the exciting and unique possibilities afforded by having a traditionally trained musician(s) *interpret* AI-generated material. There has also been a lack of easy-to-use instrument-like AI tools for musicians who are interested but have no experience in using computers in their creative process, though this has been improving recently.

This paper is a demonstration of the orchestra as a rich and exciting space in which to explore the possibilities of AI within the creative process. By continuing to place AI in an embodied, performing space, I hope we may continue to learn how humans might interact with machines, now and in the future.

## Appendix – Technical Details of PRiSM-SampleRNN Training

The PRiSM-SampleRNN models were trained on the following machine:

OS	Ubuntu 18.04 LTS
RAM	128GB
Storage	1TB SSD
Secondary Storage	8TB HDD
CPU	Intel I9
GPU	2x NVIDIA GeForce RTX 3090
Cooling	Liquid

The dataset was approximately 2000 hours of 441000Hz WAV files. PRiSM-SampleRNN trained for 7 epochs, each of which took approximately 38 hours. Only the first 5 epochs were used to generate audio.



## References

- Agostinelli A., Denk T. I., Borsos Z., Engel J., Verzetti M., Caillon A., Huang Q., Jansen A., Roberts A., Tagliasacchi M., Sharifi M., Zeghidour N., & Frank C. (2023). “MusicLM: Generating Music from Text.” ArXiv. *arXiv:2301.11325 [cs.SD]*
- André, N. A., Bryan, K. M., & Saylor, E. (2012). *Blackness in opera*. University of Illinois Press.
- Botha, J., & Pieterse, H. (2020). Fake News and Deepfakes: A Dangerous Threat for 21st Century Information Security. In B. K. Payne & H. Wu (Eds.), *ICCWS 2020 15th International Conference on Cyber Warfare and Security*, pp. 57–66.
- Caillon, A., & Esling, P. (2021). “RAVE: A variational autoencoder for fast and high-quality neural audio synthesis.” ArXiv. *arXiv:2111.05011 [cs.LG]*
- Campagna, F. (2018). *Technic and Magic*. Bloomsbury Academic. <https://doi.org/10.5040/9781350044005>
- Carykh, July 5<sup>th</sup> 2017, *AI Evolves to Compose 3 Hours of Jazz!* [video], YouTube, <https://youtu.be/nA3YOFUCn4U>
- Chan, C., Ginosar, S., Zhou, T., & Efros, A. A. (2019). Everybody Dance Now. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 5933–5942.
- Criado Pérez, C. (2020). *Invisible Women: Exposing Data Bias in a World Designed for Men*. Penguin.
- Dastin, J. (2022). Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women. In K. Martin (Ed.), *Ethics of Data and Analytics* (pp. 296–299). Auerbach Publications. <https://doi.org/10.1201/9781003278290-44>
- Dhariwal P., Jun H., Payne C., Wook Kim J., Radford A., Sutskever I. 2020. *Jukebox: A Generative Model for Music*. ArXiv. *arXiv:2005.00341v1 [eess.AS]*
- Donne, Women In Music, 2019-2020 - *Donne, Women in Music Research* (Donne 2020), <https://donne-uk.org/2019-2020/>
- du Sautoy, M. (2020). *The Creativity Code: Art and Innovation in the Age of AI*. Fourth Estate.
- Edidin, A. (1998). Playing Bach His Way: Historical Authenticity, Personal Authenticity, and the Performance of Classical Music. *Journal of Aesthetic Education*, 32(4), pg 79. <https://doi.org/10.2307/3333387>
- Engel, J., Hantrakul, L., Gu, C., & Roberts, A. (2020). *DDSP: Differentiable Digital Signal Processing*. ArXiv. <https://doi.org/10.48550/arxiv.2001.04643>

Fang, A., Liu, A., Seetharaman, P., & Pardo, B. (2020). *Bach or Mock? A Grading Function for Chorales in the Style of J. S. Bach*. ArXiv. ArXiv:2006.13329 [Cs.SD]

Fiebrink, R., & Caramiaux, B. (2018). The Machine Learning Algorithm as Creative Musical Tool. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 181–208). Oxford University Press. <https://doi.org/10.1093/OXFORDHB/9780190226992.013.23>

Fiebrink R. & Cook P. R. (2010). The Wekinator: a system for real-time, interactive machine learning in music. *Proceedings of The Eleventh International Society for Music Information Retrieval Conference*.

Fraser, J. (2019). *The voice that calls the hand to write: exploring the adventure of agency and authorship within collaborative partnerships*. Retrieved August 2020, from <https://www.julietfraser.co.uk/papers/>

Goodyer, J. (2021, October 14). *How an artificial intelligence finished Beethoven's last symphony*. BBC Science Focus Magazine. <https://www.sciencefocus.com/news/ai-beethovens-symphony/>

Gotham, M. (2014). Coherence in Concert Programming: A View from the U.K. *IRASM*, 45(2), pp. 293–309

Hadjeres, G., Pachet, F., & Nielsen, F. (2017). DeepBach: a Steerable Model for Bach Chorales Generation. *Proceedings of the 34th International Conference on Machine Learning*.

Hallström, E., Mossmyr, S., Sturm, B., Vegeborn, V., Wedin, J. 2019. 'From Jigs and Reels to Schottisar och Polskor: Generating Scandinavian-like Folk Music with Deep Recurrent Networks'. In *Sound and Music Computing 2019*.

Kanga, Z. (2014). *Inside the Collaborative Process: Realising New Works for Solo Piano* (Doctoral dissertation, Royal Academy of Music, UK)

Ma, B. (2021) *On dialogues between sound and performance physicality: Compositional Experimentation, Embodiment, and Placement of the Self*. (Doctoral dissertation, The Royal Northern College of Music in collaboration with Manchester Metropolitan University, UK)

Magnusson, T., & McLean, A. (2018). Performing with Patterns of Time. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 245–266). Oxford University Press. <https://doi.org/10.1093/OXFORDHB/9780190226992.013.21>

Mehri S., Kumar K., Gulrajani I., Kumar R., Jain S., Sotelo J., Courville A., Bengio Y. 2017. *SampleRNN: An Unconditional End-To-End Neural Audio Generation Model*. ArXiv. arXiv:1612.07837v2 [cs.SD]

Melen, C. (2020). *PRiSM SampleRNN*. RNCM PRiSM. <https://www.rncm.ac.uk/research/research-centres-rncm/prism/prism-collaborations/prism-samplernn/>

Miller, C. (2018). *Transformative Mimicry: Composition as Embodied Practice in Recent Works* (Doctoral dissertation, University of Huddersfield, UK)

Norrington, R. (2004). The sound orchestras make. *Early Music*, 32(1), pp. 2–6.

<https://doi.org/10.1093/EARLYJ/32.1.2>

Lewis, G. E. "Is Our Machines Learning Yet?" Machine Learning's Challenge to Improvisation and the Aesthetic," in *Machinic Assemblages of Desire: Deleuze and Artistic Research 3*, edited by Paulo de Assis and Paolo Giudici, pp. 115-128 (Leuven: Leuven University Press, 2021), <https://muse.jhu.edu/book/82127>.

Lewis, G. E. (2018). Why Do We Want Our Computers to Improvise? In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 123–130). Oxford University Press.

<https://doi.org/10.1093/OXFORDHB/9780190226992.013.29>

Lewis, G. E. (2000). "Too Many Notes: Complexity and Culture in Voyager". *Leonardo Music Journal*, Volume 10, pp. 33-39

Pachet, F. (2003). The Continuator: Musical Interaction with Style. *International Journal of Phytoremediation*, 21(1), pp. 333–341. <https://doi.org/10.1076/JNMR.32.3.333.16861>

Payne, C. (2019). "MuseNet." *OpenAI*.

Rohrhuber, J. (2018). Algorithmic Music and the Philosophy of Time. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 17–40). Oxford University Press.

<https://doi.org/10.1093/OXFORDHB/9780190226992.013.1>

Sargeant, A. (2017). The Undeath of Cinema. *The New Atlantis*.

Spiegel, L. (1981, January). Manipulations of Musical Patterns. *Proceedings of the Symposium on Small Computers and the Arts*.

[https://www.researchgate.net/publication/266316606\\_Manipulations\\_of\\_Musical\\_Patterns](https://www.researchgate.net/publication/266316606_Manipulations_of_Musical_Patterns)

Torrence, Jennifer (2018). Rethinking the Performer: Towards a Devising Performance Practice. *VIS – Nordic Journal for Artistic Research*

Vickers, B., & Allado-McDowell, K. (2021). *Atlas of Anomalous AI* (B. Vickers & K. Allado-McDowell, Eds.). Ignota Books.

Wang, P., Angarita, R., & Renna, I. (2018). Is this the Era of Misinformation yet: Combining Social Bots and Fake News to Deceive the Masses. *The Web Conference 2018 - Companion of the World Wide Web Conference, WWW 2018*, pp. 1557–1561. <https://doi.org/10.1145/3184558.3191610>

Whorley, R. P., & Laney, R. (2021). Generating Subjects for Pieces in the Style of Bach's Two-Part Inventions. *Proceedings of the 2020 Joint Conference on AI Music Creativity*. <https://doi.org/10.30746/978-91-519-5560-5>

Wiggins, G., & Forth, J. (2018). Computational Creativity and Live Algorithms. In A. McLean & R. Dean (Eds.), *The Oxford Handbook of Algorithmic Music* (pp. 267–292). Oxford University Press.  
<https://doi.org/10.1093/OXFORDHB/9780190226992.013.19>

## Musical Works Referenced

Herndon, Holly (2018). 'PROTO'

Lewis, George. E. (2021). 'Minds in Flux'

Lewis, George. E. (2004). 'Virtual Concerto'

Luther Adams, John (2007). 'Dark Waves'

Ma, Bofan (2022). 'That's What They Said'

Oliver, Benjamin (2023). 'Love Letter [WITH AI]'

Saariaho, Kaija (1989). 'Du cristal'

Salem, Sam (2021). 'This is Fine'

Steen-Andersen, Simon (2010). 'Double Up'

Walshe, Jennifer & Atken, Memo (2018). 'ULTRACHUNK'