On the Music(ai)lly Possible

Oded Ben-Tal
Department of Performing Arts
Kingston University
o.ben-tal@kingston.ac.uk

Abstract

This is a position paper that aims to stimulate thinking about the musical implications of deep learning tool. It proposes the use of thought experiments as a method for examining the influence of music deep learning on the music created. These demonstrate that current tools only have a limited impact which is at odds with some of the utopian pronouncements of AI-prophets. We are yet to see transformative practices of the type that emerged out of previous technological innovation such as sound recording or photography. The paper suggests that for the most part current deep-learning bases tools facilitate existing modes of music making and, therefore, exert indirect influence on the music made. The paper questions the utility of conversation-oriented textual interfaces for music generation and points to some open questions.

1 INTRODUCTION

One suspects that 2026 may be remembered as the year of 'peak AI'. That vast wealth will have evaporated in the largest economic bubble the world has ever known[1]. It is, therefore, an opportune moment to engage in some safer speculations: posing some thought experiments and entertain some 'what if…' scenarios in order to cut through some of the hype surrounding AI. A major obstacle we face when trying to discuss music AI is the lack of good definitions for music, intelligence, creativity, originality, and other terms that inevitably crop up in discussions. This paper will argue that we can clarify thinking around the meaning and significance of current AI tools through the method of thought experiments. The focus will be on the recent turn in AI research towards data-driven, deep learning methods; and in particular on the impact of Large Language Model (LLM) based tools. Fir the sake of brevity these will be referred henceforth as AI, unless specific distinctions are required.

Even after discounting some of the self-serving hype generated by tech companies, it is clear that we witnessed a transformation in AI technology in the last decade or so. Predictions about AI emerging out of specialist labs (Camurri, 1990; Roads, 1985) finally became reality when deep learning systems started generating musically plausible outputs. Using current AI-tools for music can take many forms, and if we want to evaluate the impact of this technology now and guess its future trajectory, we need to distinguish between different use cases. Typical cases include:

- Using AI-enabled plugins in the production of music (e.g. denoising, eq)
- Transcribing a recording using deep learning based systems. Using the output as material for composition.
- Using a generative AI to sketch or produce a demo version; which is then realised using other means.
- Interacting with a generative AI system (prompt engineering) to generate source material for a composition.
- Curating from generative outputs to assemble a soundtrack.
- Turning to AI powered systems for writing lyrics.

This paper is focused on musical questions putting aside questions relating to the economic impact, legal implication, and moral/ethical dimensions (Drott, 2021; Morreale, 2021; Sturm et al., 2019). We may try to quantify or evaluate the role of the AI system in each case in relation to the meaning of the music, or the creative contribution made by an AI system (perhaps in comparison to human contribution). But it is clear that these question are tangled with implicit understanding of music, creativity, and musical meaning – all rather difficult questions in themselves. Discussion will be coloured by our notions of creativity, how much attention we attach to creative process and the resulting product and what are the values we attach to music. There have been proposals to assign a tiered system of attributions to musical artefacts (Ding, 2025; Jon, 2025). Perhaps these are practical as tools for clarifying legal questions of ownership and rights, but they do not address interesting musical questions. A more nuanced approach to the question of copyright is proposed by Militsyna (2023). She devises a set of tests to evaluate human creative contribution starting with the technical dimension of the AI system and the human decisions involved in developing it; Quantitative and qualitative decision making by humans as part of the creative process; What she terms the creative distance between those decisions and the desired expression; And finally,

[1] See, for example https://broligarchy.substack.com/p/the-great-ai-bubble and another take for the sake of balance: https://www.dbresearch.com/PROD/RI-PROD/PDFVIEWER.calias?pdfViewerPdfUrl=PROD0000000000604456&rwnode=REPORT

an originality test – to what degree did the creative decisions made by humans along the way are sufficient to justify claims of originality. She notes that some of those (e.g. meeting an originality threshold) are not new problems but have existed in determinations of copyright before AI out of labs an into the realm of commercial products. Interestingly, she proposes that we will see authorless outputs – artefacts where we cannot attribute sufficient creativity to anyone. I am less interested in the legal question of ownership and copyright. Instead, I would like to stimulate musical thinking and contend that one fruitful approach is to undertake some thought experiments, starting with asking some 'what if …' questions.

2 WHAT IF ?

The first set of hypotheticals I want to explore is 'what if we replace x with y'. A non-exhaustive list includes:

- What if we replace one AI system with a different one?
- What if we (try to) use non-AI based technology for the same task?
- What if a different person uses the same AI system?
- What if we replace the data used to train the AI system with different data?
- What if we replace digital computers with human digits?

Covering each of these for each of the use cases listed previously will be laborious to perform as well as tedious to read, so I will focus on some of these to illuminate specific issues.

We can note some instances where the answer to the 'what if…?' question would be 'not much difference'. For example, when we apply standard processing such as denoising, pitch correction, or dynamic compression. Whether the tool used was developed from an audio-engineering knowledge or using deep learning, the difference will be small. Perhaps some improvement in denoising will turn a recording considered unusable to one we can use. Conversely, perhaps an unusual effect of extreme pitch-correction – which can be utilised creatively – will no longer be available with a deep learning based approach (because the mechanism becomes opaque and less controllable). But these are outliers.

Similarly, some, systems offer users so little control that the answer to the question 'what if we replace one user with another' is also 'not much difference'. Consider google's piano genie[2] where users control the timing of notes but only indirectly the pitch. It seems the aim is to produce a system that will produce an 'acceptable' pitch sequence in most circumstances. This tool seems almost musically impervious to replacing one human user by another; Or with an automated input mechanism for that matter.

2.1 Drafts and demos – AI as a facilitator

When AI tools are used in drafting process or to create demos, these tools can best be described as facilitators. Creating audible drafts helps composers externalise, and therefore test, ideas. This can include testing different alternatives, refining vague initial ideas into more concrete viable ones, etc. Demos usually serve a related but also somewhat different aim: to communicate ideas to others. These could be collaborators (e.g. performers) or gatekeepers.

In both case (drafts and demos) AI tools may open access to people with fewer skills or resources. For example, someone without access to good recording equipment or a high quality sound library, would be able to create a demo or draft with good sound quality. Similarly, users who lack knowledge (or confidence) with common practice harmony or instrumental techniques could sketch ideas with the help of AI driven systems. Note that the availability of such tools for creating drafts and demos does not, in itself, mean more people will be able to become musicians. Furthermore, while we are conditioned, culturally, to consider wider access as a positive development, this is not self-evident. But unpacking both of these questions is beyond the scope of this paper.

Creating drafts or demos is not a simple process of rendering ideas which are already fully formed in the composer's head into audio-files. The ideas are, usually, sketchy and gradually come into focus through the process of drafting (or creating a demo). Inevitably, the tools themselves will channel the composer in certain directions. In other words, the use of AI tools is not a neutral or cost free choice. Their complexity is the source of their usefulness (in this case), but this usefulness is attached to biases. For example, biases towards existing 'solutions' to creative problems. A good guess is that less experienced composers will be influenced more by the tools, while more experienced composers will be able to exert more control over the drafting process.

3 A CAGEAN DETOUR

In the 1950s John Cage composed several pieces for magnetic tape including his *Williams Mix*. Realising this piece requires recording a large number of sounds, according to 6 categories - A (city sounds), B (country sounds), C (electronic sounds), D (manually produced sounds), E (wind produced sounds), and F ("small" sounds, which need to be amplified). The process is laborious – recording hundreds of sounds, making over 3000 tape-cuts in specified shapes, applying filtering and processing, and mixing the 8-tape strips down to stereo as specified in the graphic score. Previous work investigated the capacity to translate or recreate these analogue pieces using digital tools (Austin, 2004; Erbe, 2016; Gurevich, 2015). What if we try to use generative AI to realise *Williams Mix*?

Not surprising, Gemini[3] could easily answer questions about this piece and not just simple facts. For example, when asked what would be the creative decisions entailed in realising it, Gemini pointed towards the sourcing of sounds, including the question whether it is more appropriate to try and use 1950s city sounds or contemporary ones. Another response related to the different methods of cutting tape and their possible

[2] https://magenta.withgoogle.com/pianogenie

[3] https://gemini.google.com

translations into the digital domain. Gemini was also able to suggest relevant solutions to these problem. Gemini offered to describe 'its' version of the piece and was able to justify choices – based on some of Cage's writing – when asked to explain those.

When it comes to actually realising the piece, AI systems face two obvious obstacles: sourcing sound recordings and access to the score. The score consists of 192 pages graphically representing the cutting of the 8 tape strips. Cage envisioned the strips of tape laid out lengthwise over the score and the cuts performed directly on the tape following the marks as templates. A digital realisation would require translating the marks into instructions, based on the inch-to-second ratio of the tape, the shapes of the cuts, and additional instructions (Erbe, 2016).

A second obstacle is access to the necessary sound sources. In addition to the descriptive categories, Cage also asks to limit repetition of pitched material, but allows repetition for non-pitched material. The reliability of pitch detection algorithms would, therefore, impact the success of an AI system to realise the piece. Cage also suggests the use of filtering and, possibly, other sound processing of the recordings, without specific instructions.

When prompted, Gemini provided python code that assembles what it considered the opening 30 seconds of the piece, with instructions where to place the required sound recordings. Looking at samples of the opening page available online, the code looks a reasonable rendition.

3.1 Two step generation

When asked how would it go about realising Cage's *Williams Mix*, Gemini did not provide what the author considers an obvious, AI-based solution: use a text-to-music generator. But when asked to write a prompt to plug into such a system to generate the opening of the piece, Gemini obliged. The suggested prompt looks promising:

"A 60-second avant-garde Musique Concrète composition in the style of John Cage's Williams Mix. Octophonic density collapsed to stereo. The piece must feature zero melody, zero rhythmic pulse, and zero tonal center.

Structure: Constant, rapid-fire jump-cuts every 50ms to 500ms. Use 8 layers of disparate sound categories simultaneously.

Sound Palette: > 1. **Urban**: Harsh pneumatic brakes, jackhammers, and distant subway screeches.

2. **Biological**: Close-miked human whistling, sudden laughter, and staccato vocal fragments.

3. **Nature**: Thunderclaps, heavy rain on tin, and wind howling through wires.

4. **Small/Electronic**: Amplified clock ticks, high-frequency sine wave glitches, radio static, and the sound of paper tearing.

Technique: Use extreme 'diagonal' crossfades and 'vertical' hard-cuts. Abrupt changes in volume and spatial panning. The texture should be a dense, entropic thicket of 'found sounds' where no single sound dominates for more than a second. High dynamic range, from silent gaps to sudden fortissimo bursts of white noise."

It includes relevant stylistic instructions ("avant-garde Musique Concrète"), the mechanics of rendering the piece including sound categories (but see below), and describes the overall effect ("rapid-fire jump-cuts every 50ms to 500ms" "dense, entropic thicket of 'found sounds'") .

The prompt proved a bit too long for the text-to-music generator chosen (Stable Audio 2.5), so the author had to select which parts to omit. But these choices made very little difference as none of the 1-minute versions generated resulted in even a remotely plausible result. All of them were much more sparse then the instructions suggest; Most included just two types of sounds; The sounds were much longer then expected and there was much more silence. Additionally, some of the versions did lapse into clear pulse. Stable Audio did seem to recognise thundering as a keyword resulting in thunder-like or timpani-like sounds. Sometimes a mashup of the two. Vocal sound – mentioned in the prompt – were another frequent type used.

Repeating this with Microsoft's Copilot (asking it to create a prompt for generating the beginning of Williams Mix) produced much the same result. Copilot did include the caveat that the aim is to produce something in the style of Cage, "without copying the original work" – whatever that might mean in this particular case. The prompt Copilot thinks would work is broadly similar to the output Gemini produced, but is even more obviously targeted at a human actor rather than a computational one. The text reads very much like a recipe rather than a prompt, i.e. required a fair amount of implicit knowledge not available to text-to-music generators. Both Gemini and Copilot included description of the sound categories to be used, but these were not identical to Cage's own definitions. For instance, Copilot included 'Small household sounds' and ' "Indeterminate" or ambiguous noises' while Gemini's prompt listed 'Biological'.

It is hardly surprising that Stable Audio was not able to produce anything remotely like the intended piece. After all, the music falls outside the range of outputs a typical user will want to produce and "avant-garde Musique Concrète" is, in all likelihood, a very small proportion of the training data. Nevertheless, the disconnect between the prompt Gemini provided and the resulting sounds is interesting. The text speaks to us, as human users, and aligns with how we might describe a realisations of this piece. But, evidently, this fails to produce the desired results. Describing music or sounds with words is a known challenge with a long scholarly history (Flowers, 2002; Gatewood, 1927). However, in this instance the failure is deeper – Cage's own instructions are under-specified in some respects, but what Stable Audio produced was far removed from even a very liberal interpretation. And is inconsistent with a common sense reading of the prompt.

3.2 Knowledge about music and knowledge of music

The responses Gemini provided about *Williams Mix* provided relevant information and would be indicative of knowledge *about* the piece from a human interlocutor. But it is also apparent, and not just from this particular trial, that this knowledge about music is not linked to knowledge *of* music. This distinction echoes Elliot's critique of traditional music education that treats musical works as objects for contemplation and, therefore, prioritises knowledge about music (Elliott, 1991). This approach – musical works as objects for contemplation - is embedded into much writing about music such as textbooks, music journalism, programme notes, and others. The likely sources for the responses generated by Gemini, Co-Pilot and similar systems.

The disconnect between such knowledge and the inability to generate even a ballpark approximation of the intended result can be down to two factors. One option is a type of miscommunication between Gemini and Stable-Audio, the other is down to the way Stable Audio links text and music. Both explanations can be true simultaneously, of course. Regarding the first option, both Gemini and Stable Audio are designed to interact with humans via a textual interface. Perhaps the way each processes linguistic constructs is sufficiently different such that plugging the output of one into the other does not work as expected. Obviously, this is just one example, possibly an outlier since the target music is such a niche pursuit. Nevertheless, we are too familiar with compatibility issues in digital systems more generally to suspect that this is not a unique case. Investigating how to address such short-circuiting – e.g. bridging between different AI systems such that the result is consistent with human expectations – would seem to be worth future research.

Overcoming the second possible source of error – the way text-to-music models link text and music – is also a difficult one. It was somewhat surprising that the explicit instructions "zero melody, zero rhythmic pulse" did not have the desired effect. Further, unscientific, experiments suggest that other variants of 'negative x' don't work either: e.g. no guitar(s), don't use guitars, avoid guitars all resulted in prominent guitar parts. Note that guitar heavy pop/rock music is the mainstay of commercial text-to-music systems.

We acquire knowledge *of* music through a variety of activities but primarily listening and making music (mostly of the amateur kind). Might we equip an AI system with digital ears and a digital larynx so it can hum tunes to itself as it 'reads' about music? Considering the growing recognition of the embodied nature of our engagement with music (Schiavio & Van Der Schyff 2024), focusing on audio channel alone may not be sufficient. A more productive approach may be to consider the combined human-AI interaction and find better ways of integrating our knowledge *of* music with the capacity of AI to collect and present to us knowledge *about* music.

4 MUSICAL (IM)POSSIBILITIES

So far, our hypothetical scenarios interrogated some aspects of out-sourcing portions of the creative process to AI tools. We have yet to encounter a completely novel scenario that is only possible with the use of AI. At first glace the music of Dadabots[4] fits the bill – an endless stream of, for instance, Death Metal, generated continuously. No human composer or band is capable of achieving that result. However, what about the option of replacing the generative AI with other technologies?

We can imagine the following scenario: commission ten (or a hundred) people to compose one hour of Death Metal each. We stipulate some conditions on the beginning and end of each song to enable seamless transitions. Use recombination mechanics (prevalent in video games sound) to make those into a continuous stream. Would that be recognisible different to the stream posted online? Can we even be certain that the URL for the Relentless Doppelganger[5] is not actually a very long loop? Dadabots characterise mischief as the motivation for their work – isn't a simple loop masquerading as cutting edge AI a suitably mischievous act? This is not to minimise the novelty (and savvy marketing) of Dadabots – as far as we know, no one thought of doing style pastiche on such a scale before. In other words, the generative capacity of deep learning systems inspired new, creative thinking about music. Even though the actual outcome is not contingent on generative AI technology per se. In a similar vain, creating music beyond human scale could be an example unachievable without AI. Composing a month-long piece consisting of 440 individual parts is theoretically possible using AI tools but would be prohibitively expensive without those. But the value of such an endeavour is questionable at best.

Unlike photography which gave rise to a new art form and induced radical changes in the practice of painting, or the novel musical practices that came about through the medium of sound recording, we are yet to see truly transformative use of AI in the domain of music. But the technology is still very young and changing rapidly, probably too rapidly for transformative practices to emerge.

5 SUMMARY

This paper engaged in speculative thinking to consider the musical impact of current deep learning systems. The focus is deliberately and exclusively on music making putting to the side questions about musicians livelihood, careers, and training, legal and ethical questions, or economic and environmental impact. The paper also did not attempt to explore the effect of AI (and other data-driven algorithms) on the consumption and dissemination of music.

The hype surrounding AI (one suspects that in some instances the label 'AI' is used for marketing more then substance) makes it difficult to ascertain what might be the real impact of the technological innovation. The methodology of thought experiments ('what if …') enables us to clear this fog, at least partially. The conclusion is that, for the most part, the musical impact is limited and indirect. Like any other tool, they shape how we use them which necessarily impacts the result. DAW architecture channels users towards certain ways of creating music (and thinking about music). The same is true of (Western) music notation or the logic of specific programming languages (e.g. Max vis C-sound). Similarly, the affordances of AI-based tools will influence the music people make with them. Additionally, AI tools will exert influence on creative outcomes through workflows – the way they may be

[4] https://dadabots.com

[5] https://www.youtube.com/watch?v=JF2p0Hlg_5U

integrated into existing workflows and by creating new workflows.

We are yet to see any radical new developments analogous to previous technological revolutions such as the emergence of turn-tabling as a performance or the recording studio as a tool for music making. Perhaps as White (2025) argues, music poses a particular challenge for deep learning methods. Alternatively, the technology is just too new and in time, and with enough creative experimentation, forms of music making that exploit the unique capabilities and limitation of deep learning systems will emerge.

One under explored research direction concerns the gap between knowledge *about* music and knowledge *of* music that is evident in many current music deep learning models. Augmenting text prompts with music terminology may offer some inroads towards filling this gap (Melechovsky et al., 2024). However, it is worth asking whether the knowledge required to utilise such terminology in a musically meaningful way has a sustainable future. Not to mention that many potential users already posses little or no music theory knowledge and that Western music theory is only relevant to a particular musical practice. I would argue that a better approach would be to either find alternative methods for encoding music into data and developing better ways of controlling generative models that shift away from verbal descriptions. Or, as suggested above, to design tools for human-AI partnership that optimise the collaboration through better ‘division of labour’.

Finally, I doubt that text-based, conversation mode is an effective interface for music making. The popularity of chat-based tools indicates that this mode is effective for some scenarios. Including, to an extent, generating visual material. However, most people are used to describing, in words, what they see and use verbal communication to facilitate collaboration. In contrast, most people have no reason to describe music in words (Tanner & Budd 1985). When trying to communicate about a song people are more likely to hum bits of it or use the lyrics, the band, or the title. They are unlikely to try and refer to it using music theory terms or sound describing words. In other words, prompting LLM for information, ideas, or images draws on existing modes of social interaction while text-to-music systems do not.

AUTHOR DECLARATIONS

Except as discussed in the text (mainly in section 3), AI tools were not used in the writing of this paper.

REFERENCES

Austin, L. (2004). John Cage’s Williams Mix (1951–3): The Restoration and New Realizations of and Variations on the First Octophonic, Surround-Sound Tape Composition. In P. Hall & F. Sallis (Eds), A Handbook to Twentieth-Century Musical Sketches. Cambridge University Press.

Camurri, A. (1990). On the role of artificial intelligence in music research. Interface, 19(2–3), 219–248. https://doi.org/10.1080/09298219008570568

Ding, K. (2025). A Tiered Framework for Copyright Ownership in AI-Generated Content. INNO-PRESS: Journal of Emerging Applied AI, 1(9). https://www.inno-press.com/index.php/JAAI/article/view/52

Drott, E. (2021). Copyright, compensation, and commons in the music AI industry. Creative Industries Journal, 14(2), 190–207. https://doi.org/10.1080/17510694.2020.1839702

Elliott, D. J. (1991). Music as Knowledge. Journal of Aesthetic Education, 25(3), 21–40. https://doi.org/10.2307/3332993

Erbe, T. (2016). The Computer Realization of John Cage’s Williams Mix. ICMC. https://www.academia.edu/download/48389690/ErbeWilliamsMixRealizationCamera.pdf

Flowers, P. J. (2002). What Was That?—Talking About What We Hear in Music. Update: Applications of Research in Music Education, 21(2), 42–51. https://doi.org/10.1177/87551233020210020501

Gatewood, E. L. (1927). A Study in the Use of Similes for Describing Music and its Effects. In The Effects of Music. Routledge.

Gurevich, M. (2015). Interacting with Cage: Realising classic electronic works with contemporary technologies. Organised Sound, 20(3), 290–299.

Jon, W. (2025). Prompting Creativity: Tiered Approach to Copyright Protection for AI-Generated Content in the Digital Age. Media and Communication, 13. https://www.cogitatiopress.com/mediaandcommunication/article/viewFile/9420/4370

Melechovsky, J., Guo, Z., Ghosal, D., Majumder, N., Herremans, D., & Poria, S. (2024, June). Mustango: Toward controllable text-to-music generation. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) (pp. 8293-8316).

Militsyna, K. (2023). Human Creative Contribution to AI-Based Output – One Just Can(’t) Get Enough. GRUR International, 72(10), 939–949. https://doi.org/10.1093/grurint/ikad075

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). https://doi.org/10.5334/tismir.86

Roads, C. (1985). Research in music and artificial intelligence. ACM Computing Surveys, 17(2), 163–190. https://doi.org/10.1145/4468.4469

Schiavio, A., & Van Der Schyff, D. (2024). 4E Music Cognition in Theory and Practice. In A. Boguvic, R. Timmers, and S. Nikolic (Eds) Psychological Perspectives on Musical Experiences and Skills: Research in the Western Balkans and Western Europe , 189-210.

Sturm, B., Iglesias, M., Ben-Tal, O., Miron, M., & Gómez, E. (2019). Artificial intelligence and music: Open questions of copyright law and engineering praxis. Arts, 8(3), 115. https://doi.org/10.3390/arts8030115

Tanner, M., & Budd, M. (1985). Understanding music. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 59, 215-248.

White, C. W. (2025). The AI Music Problem: Why deep learning Conflicts with Musical Creativity. Taylor & Francis.