The Conceptual Impossibility of AI Music

Sean Tartaglia

I will begin with a simple caveat, I am not really threatened or care that much about AI music, I actually am more worried about my day job because companies are stupid and will rush AI into every field for no good reason. Compositionally, I am niche enough to not even be worried about anything (I haven't even tried to get recordings for the works I have had performed, it doesn't interest me, probably a career mistake, but who cares), and I don't think the AI music is really that interesting. I am not afraid of AI as a procedurally generative process, as I have always had an interest in Total Serialism and the work of John Cage, especially his later work which is highly composed by procedural means (via a bespoke program made for him, which was, under his supervision, used to generate material for his number pieces), and I originally did write quite of lot of terrible serial music in college. In this I want to discuss some issues I have with AI music, not on a philosophical level, less so on a technical one (I am a hobbyist programmer, I don't admit to know anything), but more on a conceptual one.

As far as I understand, generative processes in music generally work in one of two ways. The first, and one I have a bit of experience in, having written small batch processes and programs for it, is procedural. Procedurally generative and modular composition has a rich history going back to early modernism (see Cowell's *Mosaic Quartet*, two decades before Earle Brown), though it is actually something I find to be generally useless, but I will explain that later.

Second is the newer form of generative process, which we see in AI. This is rather new to a degree, though generative MIDI had existed to some degree in transforming picture data into MIDI, but I think that is more a slight twist on procedurally generated music composition. To be frank, I don't find either technique to make "effective" music, but at the very least, the former can provide completely new material for compositional processes via a series of rules, whereas the second requires inputs of legally dubious origin and also takes hours and hours of training on music. (A twelve year old could use it to create "copyright free" drum breaks for his garbage

jungle track, but it is not really "copyright free" when the AI broke copyright law to make it.)

Of course, ElevenLabs did the work for us, so I'd like to muse on the way I think I understand how the deep learning model parses the music. In a procedurally generated situation, we give a low level program a series of inputs and parameters, and it uses these, to the best of its ability, to create a composition, though whether or not we can consider it coherent is up to our tastes and ultimate aims. As far as I can tell, AI Music does not necessarily have parameters hard programmed, though my only reason for this statement is because I believe the people who program the AI are generally far too shortsighted and, despite being intelligent enough to program deep learning, deeply lack any sense of self-awareness; that is, I don't think they necessarily have enough knowledge of the music to specify any major parameters or even just basic categories, since I assume it is also a lot of work to do that and they're trying to monetize around Q4 as fast as possible. They could hire a musicologist as an advisor, I still don't think they're that self-aware and they're just trying to latch on to quick AI hype, and since I am not going to die on that hill, I will leave it at that.

So to begin with a simple example: you feed the AI hundreds to thousands of hours of bebop. How does it parse this? Does it understand the underlying harmonic structure with how the alternatively voiced chords tend to muddy up the harmonic stability of a key? It's not like bebop, or even jazz in general, uses a strictly triadic language. You tend to omit a fifth, or dimish or augment it, because it is too stable in its pure form, or you might add in a 9th or 13th for color and harmonic misdirection, etc. Jazz arrangement accentuates harmonic color just as much as harmonic movement, sometimes, if the arranger is overzealous, to its detriment, as in a *Wave* chart I once played where every chord was so permuted I was incredibly annoyed by how the harmonic movement ended up totally aimless, which did, to me, a big disservice to Jobim's melody.

So when we feed the AI thousands of hours of bebop, with all kinds of different chords, keys, voicings, progressions, how does it make sense of this; and, moreover, how does it make sense of this in the context of *other* music you have fed to it? Might this eventually leak into other music based on how prompts are written?

I am not really asking this for an answer, I don't really care about the justification for AI existing because I don't believe in infinite utopianistic progress; for me, it is much more a question of: does the AI actually understand music conceptually, not really just the theory, but the music its fed; does it understand the formal decisions that

were made, beyond verse chorus verse, enough to actually make something coherent, or is it just going to throw up incoherent or muddled progressions because it was fed something too complex?

I want to take a moment to return to the first form of procedural generation, something many programmers seem to find fascinating, as you'll never not find them going to reddit or stackoverflow to ask people about theory so they can write some low level counterpoint or common practice harmony batch file.

The problem with the programmer, who could be very intelligent, is that they immediately approach music from an incorrect conceptual level: theory comes after the music is written, and music is not written as if paint by numbers outside of some Fuxian counterpoint example. Counterpoint, in the traditional method of teaching, comes from Palestrina's music, placing emphasis on rules derived from the way he handled line and dissonance, especially in comparison to Christian chant, which something Jeppesen highlights in his study of Palestrina when discussing melodic line.

In the example of counterpoint, the rules of verticality, the harmony, and handling of dissonances via movement and suspensions, are not really the end goal, but rather, the result of melody. If you simply add in all the rules of harmonic motion, you'll end with with something that probably won't move that much, something like an Eric Whitacre choral work, because the program will be checking each instance of a harmony against the rules incredibly strictly; however, in counterpoint, the vertical harmony is not necessarily the main point, just what the main point conforms to in the Palestrinian compositional practice.

So you need to make sure to tell the program to place more emphasis on the rules of horizontal line via stepwise movement and in leaps of a third, fourth, fifth, and sixth. Now, with this in mind, the procedural program will, without a concession to taste, write a line that conforms to all of these rules, but will probably miss something in the question of taste: Where does it place suspensions, leaps of varying size, stepwise sections, dropping voices for duo or trios; essentially, the procedurally generated music is not "inspired" melodically, it is written like a counterpoint study, and, I am sorry, it is probably not that interesting.

Even in the context of common practice harmony in its most basic form, A figured bass' harmony isn't always implied by the melodic information, so the decision to use certain harmonic configurations is one made with some sort of extra-systemic intent, because there are many ways to work within a structural system without having to adhere to the rules precisely as written. If a

procedural process were to write basso continuo, it would most likely write pure triadic progressions, and not add in 7th or 6th chords, because there is no real systemic rule to write these harmonies in, unless maybe you wrote a rule that told the program to write one in randomly every 4 bars; however, this kind of misses the point. The choice of harmonic configuration is one of taste and context, a mark left by a composer that cannot be properly replicated by following systematic rules.

The reason why this is important to explain is that this attempt at generating musical material had been done in serialism—and still is being used to a degree, as in Ferneyhough's compositional process (Ferneyhough uses a bespoke program made for him that generates complex rhythmic material for him to arrange to his liking)—where parameters are highly structured according to various rules and inputs, sometimes employing the low level computation of 50s. The problem is, a majority of the composers associated with Serialism in its most rigid form essentially exhausted its compositional possibilities by the end of the 50s, because at some point you can only serialize so much before the compositional process becomes intellectually and creatively unstimulating. As far as I am concered, Cage's *Book of Changes* basically takes both Serialism and Chance to their limits, and that was written in 51. Stockhausen moreso, since he exhausted everything you could do with the gigantic budget of of European radio orchestras in the 20th century. Every original total serialist eventually moved on the more "post-modern" music in the 60s, what we call the "Post-Serial Thaw," aside from Xenakis, Babbitt, and to some degree Wuorinen, though each had a more "expressionistic" quality to their music not found in the 50s (Xenakis wrote highly abstract computer music, but when he wrote for voice it was much more dramatic than general Serial work).

AI music comes to this from the other side of the conceptual conundrum: where procedurally generated music totally understands rules and parameters, but lacks intuition and inspriation, AI music, being fed on music, has an understanding of general melodic and harmonic patterns, but doesn't necessarily seem to understand why. Note that when ElevenLabs posted the "jazz" music, it was essentially Michael Bublé. No offense, but that's crooner music, and it's not jazz, it's pop music played by a big band. AI music comes to generative processes from the top down, it notices style and effect, without an understanding of why, which is why everyone who listens to that kind of slop tends to be an idiot who is only interested in music as a sort of surface sheen product, essentially everyone who makes trap; in contrast, procedurally generated music comes to generative processes

from the bottom up, it is built upon parameters and rules, but does not have the ability to understand where and why patterns in music are interesting to us, which is why everyone who composes it tends to be an idiot who is only interested in music as a sort of articifically constructed structure, like building a cathedral for the sake of it being tall, as opposed to it being also a physical metaphor for spiritual teachings and religious tenets.

We talk quite a bit about intent when it comes to the problem of AI, but there is more than just intent in artistic expression. I quote this quite often, but Schoenberg's own words still ring true, no matter what era we live in:

> *Whether one calls oneself conservative or revolutionary, whether one composes in a conventional or progressive manner, whether one tries to imitate old styles or is destined to express new ideas—whether one is a good composer or not—one must be convinced of the infallibility of one's own fantasy and one must believe in one's own inspiration. Nevertheless, the desire for conscious control of the new means and forms will arise in every artist's mind; and he will wish to know consciously the laws and rules which govern the forms which he has conceived "as in a dream."*

"Good" composers tend to come to music between these two poles, so that style itself is not of prime interest, and neither is the means to reaching it really important compositionally; rather, it is the interesting result of these things in combination with individual inspiration, of what the composer has experienced and wishes to express. When you work with music as a composer, arranger, or performer, you're grappling with many creative decisions informed by your personal practice, i.e. those things that have influenced you in your experiences of music in study and performance. What has been detrimental to our musical society, and perhaps in the whole of our artistic society, is that we live in a world where music has been so societally devalued that most children do not even recieve a musical education, and they cannot have these critical sorts of experiences that foster a love for musical exploration and investigation, instead their experiences of music are always secondhand, as an observer, a consumer, as a nobody; and that, to me, more than anything, than any AI, is what destroys music, because it leads to a complacency and lack of interest in growing and developing one's musical experience,

in both activity (playing) and passivity (listening). This denigration of one into this consumer, this nobody who exists only to make others rich, seems to me the cause of this zealous AI art movement, and those people who obssess over it, in any form, because it seems to carry this very sad, jaded mentality, this bizarre hatred of people who worked and sacrificed many years to develop unprofitable artistic skills. Trying to interact with these sorts of people is impossible, they make fun of artists when they are poor, but when they make a few hundred dollars on comissions, they get very mad at them for making money. To me, this is like complaining about people who dedicate themselves to cooking opening resturants, gloating when they fail, and wishing to have an AI chef at home cook 3 star dinners to stick it to them. To hate people who give up everything for these things, it is a totally irrational concept to me, and can only be driven by some sort of spite or malice caused by some ineptitude or insecurity. These decisions we make aren't jobs or careers, these are vocations; that is, one chooses it because it is intrinsically part of one's life, and there can be no other thing in one's life other than that thing. For me, dedication to one's craft is like becoming a monk (and this is not because I mainly love and compose liturgical music), one takes, in some form, vows to dedicate one's life to these things. I cannot begin to understand how people can live so shallowly, so hollowly. Music is like life, it is not something to consume mindlessly, as if living without reflection, but something to enjoy and reflect upon, something to live with, even if it brings hardship. In one of my oldest writings, I wrote about this way of thinking as so:

> *A musician is like one who lives in a house with music, and the music with which he lives is his family; he is born with them; he laughs with them; he suffers torment with them; he eats with them; he sleeps with them; when he is away from them he dreams of them; when they die he mourns for them; and, when he dies, they die with him.*

I am, and always will be, very Ruskinian in this sense, as for me, being a composer or musician, being with music, is no different than being a genuine person; those things you accept into your vocabulary determine the certain kind of person you are, the sort of decisions you make; the music you choose to surround yourself with determines your overall beliefs. You draw in order to see, and you compose and play in order to hear. If there is no wealth but life, the

quality of that wealth is that life you choose to live.

That aside, I am genuinely curious as to what AI music does when fed counterpoint and being asked to write a prolation canon. I know exactly what a procedural generative technique will write, it'll be very boring and very safe in order to maintain harmonic stability according to the parameters, but how does, how can, AI, when given a work by Ockeghem with 4 voices moving at different rates, with a cleverly hidden construction, parse it, and what does it spit out when asked? Is it just going to be that mixed in with other vocal music, and then, will it even be able to put out a true canon at all? If music is surface, then it shouldn't be that hard to fake Mozart, in common practice harmony it'll probably just put out some inferior version that sounds like a fourth rate classical era composer; however, if music is something more, then what happens when it is given Bruckner, where harmonic rules are constantly broken? What will it put out?

Now people believe (and they say it is already happening) that AI will just get better and better, and while that might be true, dimishing returns are real, and once the internet is filled with AI generated material, it'll probably cannibalize itself and muddy the material in the process. I think the same could be said for AI music, since Spotify decided to ruin the sample set with bland muzak. AI has the potential to be good if it has proper parameters and produces MIDI according to those things. It cannot be good if it sucks up pre-rendered audio data, because the information, as far as I am concerned, cannot be properly parsed. If it were purely trained on audio stems that were carefully categorized and given specific parameters, as in the earlier example of a breakbeat, and was able to stitch these together, then maybe it could grow; yet, that would also require compositional intent, which I am not sure it could do given the current sets of training data.

I am not interested in the argument that "this is the end of music," because the people who want to consume this slop were probably never your audience in the first place. No composer should be threatened by this unless they're a sub-par beatmaker, because AI music inherently is uncreative and formulaic, there's simply no way for it to be other than that, because it has no conceptual framework to do anything other than the formula its provided. You shouldn't be worried about AI, you should be worried about the people who claim to be human.