AI and You

Transcript

Special Episode: AI in Music

Episode 88

First Aired: Monday, February 21, 2022

Hello, and welcome to episode 88! So guess what we're going to be talking about today? Some time ago, a listener asked if we could do a show about AI and music, and so I went looking for experts and found not just one but three and here we are! **Bob Sturm** is Associate Professor of Computer Science at KTH Royal Institute of Technology in Stockholm, Sweden, and lecturer in Digital Media at the Centre for Digital Music, School of Electronic Engineering and Computer Science, Queen Mary University of London. **Dorien Herremans** is an Assistant Professor at Singapore University of Technology and Design. Before joining SUTD, she was a Marie-Curie Postdoctoral Fellow at the Centre for Digital Music at Queen Mary University of London, where she worked on a project with the simply awe-inspiring title of: "MorpheuS: Hybrid Machine Learning – Optimization techniques To Generate Structured Music Through Morphing And Fusion." And **Hendrik Vincent Koops** is an AI researcher and composer, holding degrees in Sound Design and Music Composition from the HKU University of the Arts Utrecht, and degrees in Artificial Intelligence from Utrecht University. He is currently co-organizer of the AI Song Contest.

The music you heard at the beginning was a composition that Bob generated with the AI folk-rnn, and piano accompaniment by another AI called Music Transformer. He called it "Wordle's done. Now what?" (Because he was inspired by playing the game.) It's linked from the show transcript in his blog "Tunes from the AI Frontiers."

If you were following where those folks are in the world, you'll realize that with me on the west coast of North America we spanned the globe in terms of time zones. It was 4am for me. You can tell. Fortunately, my guests more than made up for my lack of with-it-ness. Let's get into it!

Peter: Dorrien, Vincent, Bob, welcome to *AI and You*.

Bob: Thank you.

Vincent: Thank you, Peter.

Dorrien: Thank you.

Peter: So we're here to talk about AI in music. And you all have some various expertise in that, that I want to tap into to give our audience an idea of just where AI is in music at the moment and what it is doing. I remember, even some decades ago, demonstrations that AI could imitate classical composers, Bach, I think. In the intervening time, where have we come? What is the state of the art at the moment? Where are we now and where are we going?

Bob: I think that is evinced by the sheer number of companies that have been founded or and acquired by bigger companies whose mission is to generate music for games, for YouTube, TikTok videos, for commercials, production music, etc. The landscape is quite populated now across the world with these companies that are using this technology for such purposes. So within these rather limited application domains, there's been a significant enough amount of progress that the knowledge is profitable right now.

Peter: So you're telling me that it is good enough now that there are people making money on it in certain domains. You mentioned some of the applications here, and those are what I would think of as being that cheap music, ones like elevator music, "We need music here, but it's not the main purpose. We want to spend as little as possible on it and it doesn't have to be that good." Is the state of the art approximately there or is it better, and the industry hasn't caught up yet?

Bob: I would say it's there in many of the cases. I mean, you mentioned Muzak, music to be piped in for wallpaper, using Brian Eno's term for this kind of music, not for the forefront, but to accompany, say, film, video game situations. But it's a very cheap alternative to human labor. And I would say it's cheap, but it's not low quality. It's surprising quality.

Peter: Right. So now the question becomes whose labor are we replacing? We get that the person who's writing the Muzak is in trouble, but does that mean it's coming after John Williams? Dorrien, do you want to take that one?

Dorrien: Yeah. Well, I think maybe the question should be what jobs are we providing? Because there's all these young researchers in music and AI technology who are getting jobs as creator of these AI models, which is something that shouldn't be forgotten. I don't think necessarily that we're going to be replacing composers as such, because there always will be a place for human composers, people will want to have those sorts of human idols as well. When it comes to AI, I think, yes, they can be used as standalone systems, but there's a really good opportunity for them to be used as an aid to composers. I'm not saying that John Williams needs this, but another composer might benefit from having a tool that can suggest ideas or that can help him compose more quickly. It's sort of like when we're doing image editing now, our software, you can say, "Okay, select all the persons in this image." The AI tools actually give you the power to say well, "Okay, now give me a chord progression that sounds cheerful," and it can just give you that sort of higher level of control over the composition process.

Peter: That is fascinating. That is a parallel with centaur chess, where computers help humans play chess and add a dimension to it. And so you're telling me then, that the state of the art in the software is such that you can express what you want in these higher-level terms.

Dorrien: Exactly.

Peter: Make that visible for me, describe that a bit more. What does it look like if I'm composing a piece? What do I see? What are my choices?

Okay, well, there's tools that, let's say, you can have a chord progression, and then the AI tool would compose a melody on top of that, and you might be able to select, "I want it in the style of this composer or that composer." And a lot of these tools, I must say, they're currently in the form of Python code and the user interfaces, they're really still very much under development, but the core AI technology is actually there.

Bob: As another example, Jukedeck was a startup company in London and was acquired by TikTok a few years ago. Jukedeck had an application that would run on the web, where you would specify, "I would like moody piano music to last 72 seconds and I would like the climax to occur at 61 seconds," and it generates from beginning to end, a fully mastered moody piano track that meets those constraints. These are high-level terms that a machine has been supposedly trained in and at least hits close to the target of what a lot of people would think of as moody. That's the problem with a lot of machine learning, is it's using these terms that are pregnant with meaning, and it leads to unexpected results sometimes, but these are high-level terms that need to be unpacked.

Peter: So, then that sounds like some sort of cluster analysis was going on for it to have learned what "moody" was and that you could feed in some other kind of emotion and it would know what that was. Is it learning from labeled data there?

Vincent: Yes. So usually what you do when you want to use such a model like that, you would have a data set where you have music in whatever kind of form, you have it labeled with those kinds of labels, in this case, moody or other kinds of descriptions. So you would train a model in that and it would generate some output or classify - tell you, "This piece of music, this is moody" or is something else.

Peter: So is there some equivalent of ImageNet for music that has been labeled according to how it makes you feel?

Bob: There's an audio form that's been created by Philippe Pasquier's group, where you can say, "Midnight market in Bangkok," and it produces a 20-second clip. You have crickets, you have sounds of Thai language, cars maybe.

Dorrien: Very cool.

Peter: Wow. Okay. So that's demonstrating a lot, and what I'm hearing is that you've got now a big data set. So, image tagging took off when ImageNet was created. Fei-Fei Li had to do that, and until that point, you couldn't do much because deep learning requires data. If you want to compose a moody piece, then something has to have associated moody with some types of music. Is there a database of that?

Dorrien: Yeah. When my lab was working on music generation with emotions, this was definitely one of the challenges we had. Because also, music generation systems are typically not trained on audio, but on MIDI, and we just don't have that much MIDI data sets out there. But recently, there have been some data sets released, I think Fiji MIDI, there's AudioSet by Google, [which] is a list of audio with YouTube clips, that have all sorts of labels like "cannon shot." And also, I don't want to put a number on it, but it's probably around 1000 labels.

Bob: And there's the Essentia project at Music Technology Group at Pompeu Fabra in Barcelona, Spain, which has a particularly inventive approach to acquiring data. You download an application on your computer, it analyses your music audio collection, and computes features from that audio collection, and associates them with labels that people have given them at other websites, like Last.fm, or Spotify, or YouTube, or elsewhere. And so you have a massive collection, I think 10 million tracks at least at Essentia that can be used for these things. Now, the data quality, and why people choose to apply particular labels to particular pieces of music, that is oftentimes swept under the rug. Let us assume that there's signal in the noise, and the more data we have, the better opportunity our models have to learn some of these important directions and whatever space it's working in, whether it's symbolic or acoustic.

Peter: Right. And that would work for a piece that was monotone in its expression, someone could label this 30-second clip as being "moody." But a lot of music varies dramatically over its length. And people don't often go to the trouble of labeling which part is which.

Bob: No, and that's a big problem I see in research that looks or seeks to extract information from musical audio, is the assumption that a musical document - the recording itself is the music that we need to analyze. And an entire world of context is completely ignored, and out of reach of any computer. But again, we need the data, our machines, they learn from data, let's put it in and hope for the best, and then try to make sense of it afterwards.

Peter: So where's the leading edge on this exploring at the moment? Is it in finding innovative ways to label data? Or is it different ways of creating the music? I mean, there's now a field called computational creativity. And so there's not just a few people researching this, this looks like it's got to be a big field now, relatively speaking. So what's the hottest topic that people are pushing on?

Dorrien: I think there's a number, and one certainly has to do with data sets. And to create data sets, there is a very important unsolved problem, which is audio or music transcription. If you can transcribe music from the audio into the score or the notes or the MIDI representation, that will enable music researchers to do a lot more with audio data sets. On the other hand, you mentioned ImageNet before, and the advances in ImageNet with a data set were not only made due to the availability of data but also due to the architectures that were developed like ResNet, MobileNet, all these things. And audio is kind of a different problem because we have this time aspect as we have the third dimension. And there's been a lot of advances recently with WaveNet, with temporal CNNs. And I believe that finding these pre-trained representations like you have ResNet-50 for ImageNet, we need to develop these representations so we can properly represent audio. And this year at the NeurIPS conference, we organized the HEAR challenge, which was exactly this. It was asking participants to create these multi-task deep representations. So we didn't give them any data set, we just asked them to pre-train models, and then tested it on like 20 different tasks, and saw which one was most robust. So it's the first year it was organized, and I hope with this initiative that we get some people to focus on this.

Peter: And that reminds me of the AI Song Contest, which is a catchy title. And, Vincent, you're involved with that, can you tell us what your role was with that?

Vincent: Yeah, so I'm one of the organizers of the contest. So it's an international contest where teams can sign up and they can make songs with the help of AI. They submit an audio file and a document, and in the document, they describe how they use the AI as part of their song writing process. And the goal is not to replace musicians or to replace composers, it's about finding out new ways to be creative together with AI. So it's sort of like as your creative partner as you can have in a band, for example, or in another way in your song writing process.

Peter: Right. If we were to take a chart of word pairings, "song contest" would most commonly be preceded by "Eurovision." So, catchy title; reminds me of the Animal-AI Olympics, where there was a contest to replicate the intelligence of different levels of animals, see if you can get up to the level of toad, for instance. So what did the entries into that contest demonstrate? Any surprises?

Vincent: Yeah, a lot of surprises. So because it's kind of like an open-ended contest, we ask that the teams submit a song, we don't really define what a song is. Usually, what is commonly understood is that it has lyrics, it's not too long, so we've restricted it to four minutes. But basically, anything is possible. And what you see there is that the teams explore that space in various really interesting ways. So lyrically, their models are exploited, or you could even say abused to generate really interesting text combinations, but also, on the musical side. So generating melodies, or chords, those kinds of things are all explored. And what you see is that there was this interesting progression from the first year into the second year, where, first of all, we doubled the amount of teams. But we also saw that the AI sort of took a more integral, interesting part into the song writing process, where in the first year, it was mostly, "We take this model, we generate something, and we are going to make a song out of it." But what we saw last year is that people actually took the output of the models, changed them to their liking, and then created new models on top of that, so it was kind of like a second-order, AI creativity, sort of like this feedback loop. Which was super interesting, because that tightly integrates the AI even more in the creative process.

Peter: We're talking a lot about "creativity" here and this conversation about "what is creativity" comes off endlessly. Has this kind of music generation expressed a surprising creativity? Has it done things that people haven't thought of? And the sort of example I'm thinking of is like AlphaGo and its famous move, which people [were] taken aback by. Of course, it was something that was in its search space, but its search space was a little bit bigger than humans at that point. So have you seen AI music generate things that have surprised musicians in, let's say, a good way, other than, "Oh, well, that's rubbish"?

Bob: In the early days of my work on this, with a system called folk-rnn - it's been trained on 27,000 traditional Irish dance tunes - we hired experts in Irish traditional music to come take a look at some 1,000 outputs of the system. And we met over a beer and he started to go through some of the tunes, and he said, "Okay, this one right here, this has a very interesting figure. It goes up in a way that I've never seen in any Irish tune that I know," and he knows hundreds of these tunes. And I wondered why not? It sounds good to my ears. It fits within the tradition, but I hadn't seen it before. And the tune itself is surprisingly good. He remarked that of that early iteration of this algorithm, that one in five were surprisingly good in his opinion.

Peter: That's exactly the sort of thing that I was looking for. And I see that there was a conference in 2020, the Joint Conference on AI Music Creativity. What spaces does a conference like that explore at the edges, other than just, "Here's our model. Here's what it's doing"? What does a conference like that generate as its collective collaborative product?

Bob: That was a conference that I organized as a kickoff of my Music at the Frontiers of Artificial Intelligence and Creativity grant. It was the first time of bringing together to communities that had been separated up until that point, one of the communities is called Musical Metacreation. And it comes from a variety of researchers, Philippe Pasquier, Arne Eigenfeldt, Ali Momeni. And it focuses on the artistic side of applying machine learning or algorithms in general, to music creation, and creating musical partners. The second community is the Conference on the Simulation of Musical Creativity, which started in 2016. And so I proposed at this particular conference, to bring together these two communities for the first time. And then we're going into the third conference this year, which will be in Japan. But the output of such conferences are research papers, musical compositions, keynote talks from not only academics but also people in the industry. So we had, from OpenAI, Christine McLeavey. She's the one that created the Jukebox, great Jukebox application. So it's a wide variety of outcomes, outputs from these kinds of conferences.

Peter: I'm just struck by how much this field is developing, and so quickly. I'd like to hear from each of you actually, how you got into this field. Was it where you started or did something lead you into it from somewhere else? Dorrien, perhaps?

Dorrien: Yeah, it was where I started with my research. It wasn't where I started with my studies because I was doing a business engineering major, so I was studying economics. And in the thesis topics, I saw this topic proposed, "metaheuristics in music," and I was really tired of thinking of profit and economics. And I was like, "I don't know what metaheuristics are, but I'm going to do this thing with music because it sounds creative and more fun." And then that's when it started. And in 2004 already, I used an optimization algorithm that they use in economics for timetabling, to generate melodies. And it's all sort of started up from there.

Peter: Vincent?

Vincent: Yeah, so I guess the first step in my academic background was studying music, actually. So I'm a composer for film and for commercials. I studied sound design and composition. And when I finished that, I didn't really see myself becoming a full-time composer. I saw myself doing that because I like writing music, of course, but I didn't really know how to be one and make a living out of that, I guess. And I was really interested in artificial intelligence, so I started studying that after finishing my master's in composition. And there, I met people at the Multimedia Research Group, who were actually combining these two things that I really liked - AI and music. So it wasn't really focused on generating music or composing music, but analyzing data sets of music. And I stayed there to do my PhD in basically music information retrieval. So music and AI, you can say, that's sort of my background there.

Peter: Thanks. Bob, what was it that grabbed you?

Bob: Well, my background is actually in digital signal processing and music. My PhD was essentially about the analytical equivalent to granular synthesis. But in 2015, when I was in London, I read a funny blog post by a computer science student at that time that was applying a simple and well-known machine learning model to simulate texts by Shakespeare or Wikipedia text, and he provided code and I said, "Oh, this would be kind of cool to apply to this large dataset of Irish traditional music that I knew of." And so just as a Saturday afternoon sort of exercise in humor, I wanted to create an endless loop of fake Irish traditional music, let's say. I was up and running within 24 hours and producing these things, and I was like, "Wait a second, this is quite surprising quality." So through these fits and struggles, I eventually ended up where I am now, where machine learning has now taken over the majority of my research endeavors and signal processing has really taken a backburner to my research.

Peter: Thanks. So there's several things mentioned there that I want to amplify. Dorrien, you were talking about how you adapted software that was for processing economic data into music. And I think it's one of the things that is striking about AI is how multipurpose it is. I just had a news item about an MIT researcher who was working in natural language processing and used the same software to analyze mammograms. And so we can have software that has been used for some completely different purpose in AI and it's relevant to what we're doing in this field. Do you see that expanding? Is there something happening right now where you think, "Oh, if I grab this thing that people are using over here in GPT-3 and put it in music, well, that could be really interesting"?

Dorrien: That's a good question. I'm not entirely sure. I think there's a lot of strong machine learning models out there that I think haven't properly been applied to the field of audio or music. And I think that's often the case. Advances seem to happen in the field of vision and NLP, just because they have a million more researchers than art. So some things, I think, that have potential are semi-supervised or unsupervised learning techniques, simply because we have to deal with such small data sets. Like we might not have a lot of emotion labels, but we have a huge collection of MP3s, so we can sort of pre-train our model, and then use some semi-supervised technique to only train a little part of the model with these emotional labels. Yeah, and other things that are very important, I think, are models like transformers like WaveNet that capture these temporal sequences a lot. And I think that's where we sort of overlap a little bit with NLP.

Peter: Well, and speaking of transformers, do you look at something like GPT-3 and get data set envy and think, "If I had a 175 billion parameter transformer, and it was trained on the EMI Catalogue…" Do you petition OpenAI to do something like that? Has anyone gone to them begging for that?

Bob: They have already, in terms of the Jukebox application. It wasn't an entire catalogue, but it was something like 600,000 tracks that they've added their autoregressive models on. It's not a GPT… it is a transformer. I'm not sure. I don't think it is. It must have attention in it somewhere. I'm not sure.

Peter: All right. 600,000, but then there's more music than that out there. Do you think, "Oh, what I could do with 60 million tracks"? Or is that not where your main desire is leading at the moment? If someone came into this field, if OpenAI came to you and said, "You know what, we just got a $50 million grant, it specifies AI music on it, what should we spend it on?" What would your answer be?

Bob: I mean, one of my answers would be on explainability. For instance, you train a model; you don't know what it's learned. There are components of that model that's responsible for some high-level musical notions, phrasing, repetition, rhythm. Where are those? Can you isolate that functionality of the network? Can we interrogate the machine to figure out what it's learned, and how to train it better? How to tune it to particular compositional voices or uses. I'm also a big fan of breaking things. So taking a model and applying it to things that it was never meant to be applied to, or performing lobotomies on it to see what happens. How does it respond to particular inputs? Find the limits of its knowledge and exploit those creatively. I think also working with artists. So use a big portion of that money to take back to the labor that's essentially being replaced by these automation engines, and bring in their expertise in order to enrich our development and application of these systems.

Dorrien: I would definitely focus on controllability as well, of the models, because I envision this model where it's generating for you on the go or live, and you have these sliders, like in a music production studio, and you say, "Oh, I want some more happiness. I want some more piano," and you can sort of control it on the go because you might want to control how the flow of the music goes over time. I think that that is very powerful to do. And you might have a big data set of 60 million songs, but you can't just train a model on rock and pop songs at the same time. These music genres have very different features so you need to have these controllability parameters as well. I think a second aspect would also be music production. Because companies like Jukedeck, who got bought by ByteDance, as Bob mentioned, they spend a lot of their efforts on not making the music sound like MIDI, but actually making it sound like it's played by a human. And that's like humanizing the timing of the performance, like making little pauses like a performer would do it. But also, producing it like it's produced by Sony or something. Yeah. So I think those two things are worth it.

Vincent: So I think it's related to what Bob and Dorrien also mentioned. So this is coming from the perspective of the AI Song Contest and working with artists. So what you see there is that often, the artists or teams start out with the idea of, "We're going to create this model, we're going to train it on this data set of songs, and it's going to generate the song for us, and we're going to just submit that." So it's going to be like one model that's basically it's going to generate a song. And nobody is able to do it. They all fail at creating this one, single, big model that's just going to generate the song, and that's because it's a very hard problem. It's very hard to generate or to create such a model that is capable of learning all these intricate parts that are needed to actually create a song. So what you see what the teams do, they actually create small models that are able to solve all these individual tasks so they can generate texts or they can generate lyrics. They can generate some melodies. They can generate chord sequences, drums, those kinds of things. But then they run into the next problem. How are you actually going to combine all those things? Because now you have all these different models that can potentially generate millions of outputs, hours of these parts, and there's no way to combine those things automatically. At least it's very hard to do. Because you have, for example, a bunch of lyrics and you have a bunch of melodies, but you need to match those to be able to find out which lyrics actually go with what melodies. And this is also, usually not the way humans make music, either. It's not like somebody writes lyrics and then a couple days later, he comes up with the melody and then checks like, "Can I match those things?" You come up with those together. Composing is doing multiple things at the same time. But we cannot do that. So that's one thing. And then the last thing is also what Dorrien also mentioned, is that, once you have a model, and it's capable of generating something interesting, what if you want to change it a little bit, you want it to sound a little bit more moody, to take the example that Bob used before? It's really hard to go into the model and find those parameters where you want to tweak it to generate sort of the same output, but slightly different. It's really hard. You basically just have this, for lack of a better term, I guess, like a black box, where you have this model, and you have to stick with it, and you just have to go with the output or change it manually.

Peter: Is that because it's hard to describe what you mean by "a little bit different," like what dimension you're talking about?

Vincent: Yes, that's definitely one problem. There's a big, I guess, what you call a semantic gap where the descriptions that we have like moody, for example, or how the computer represents that in a neural network, for example, if you trained it on those labels. It's just a way of solving that problem, matching the input to the output. It doesn't really mean if you're going to wiggle around with that parameter, that it's going to change in the same direction as we feel we experience moodiness.

Peter: Now, when you were talking about generating parts, it made me realize that I have a fundamental gap here in my understanding, so let me describe it. If I think about a piece of music that's got multiple voices in it, multiple instruments, polyphonic sounds, an orchestra, we create that by a score that's got a dozen staves on it for different instruments. And then you hear that. Now, if that's the sound that the AI is being trained on, does it somehow deconvolve all of the parts out? Or can you even train it on multi-voice recordings?

Vincent: You can.

Peter: So how does that work, then, in creating new music? Because we would create music by layering all those parts together. It's not had the experience of separating them, has it?

Vincent: So it really depends on what you use for generating or creating your model. So if you train it on audio, on sound, you basically train it on the compound sound, the full harmonic thing. And what it will try to generate is that same kind of sound, but it doesn't really know that it has multiple voices, or not explicitly like that. It's just that particular kind of sound. It's trying to mimic that kind of complex harmonic sound that you fed it. You could, of course, feed it sheet music, and there you have different layers and different voices. And then you can generate those different voices. But what you're generating then is sheet music of course and not the sound itself.

Peter: Is the way to go between sheet music and the sound established well enough for that to be—Bob, I see you shaking your head.

Bob: It is in some practices. So much of our research is focused on music of the past 300 years. Bach, Mozart, classical forms where the notation, the vertical and the horizontal are well described by this common practice notation that is so arbitrary that it's amazing when you look at it, and it ignores a lot of the modern-day music practices that are focused on tambour and transformation and spatialization. So back to this point of a system that recognizes sounds and tries to unmix them, there's been some very interesting software developed out of IRCAM by Philippe Esling, called Orchidea. You give the computer a sound and it finds a way to reproduce that sound using classical instruments. And it produces a score of what instruments play when, what pitches, what dynamics, in order to make something that sounds like cars honking, birds tweeting, a variety of things. And it's very successful. So in some aspect, the machine knows something about the ranges of instruments and their tambours, how they combine to create these amalgamations of sound to mimic this other kind of sound that is being fed to it to say "mimic this." But for all the machine knows, it could be working with stock prices, it could be working with seismic data, wave data, it has no idea about sound as we're describing it.

Peter: It's just numbers to it. But I'm tempted to think, and I wonder whether you could take the large amount of data of classical music recordings and large amount of data of classical music scores, and here you have your inputs, and say, "this equals this," in the same way that we take French and English translations and say, "these ones are equal," put them in and train a model to understand what a score was, and then give it another one and say, "play this"? Has that been done?

Vincent: So what you could use for that, for example, is a conditional sample RNN, for example. This is a model that is capable of generating raw audio. And a simple version of that is that you could train it on some data set of audio, let's say the string quartets by Beethoven, for example. And then, once it's trained well, you can generate snippets that sort of sound like a string quartet. But that's unconditional. But what you could do is you could condition it on, for example, parts of the score, or you have to make choices there [of] how you want to represent that. But then given such a part of a score, you could generate the sound as if the string quartet that you used for the sound input played that part, for example. That would be a way of doing it. But the problem with these models, I have found, personally, is that they kind of start wandering off, and it kind of sounds like they have sort of a sense of the sound of what you fed it, but it's not really capable of really understanding the detailed structure of it. It's kind of like when your grandmother would explain death metal to you. So she has an idea it's a lot of noise and a lot of drums but she couldn't really explain the details of what makes that genre that genre.

Peter: And that's one of the problems, it seems to me, in say-- Again, classical music would work for anything but if you're sampling that, then the different instruments come in and go at different times. You've got drums, and then they stop, you've got the clarinets, and then they stop, and then the trombones come in. And at any given time, half a dozen of them might be playing, but they come and go. If your AI is listening to this as just a melange of everything, it doesn't know that you've got an instrument stopping and starting with some kind of purpose, and so does its output sound as though instruments stop and start at times that are congruent, that are whatever the word is for "nice to listen to"?

Bob: Yes. I think some of the work by data bots, actually, they have 24-hour YouTube channels of death metal generated by a neural network, not only death metal, but math rock, and also jazz. And you hear structure and things coming in and out and being rephrased, and it's amazing that the network is so successful at producing the audio signal. It has nothing to do with symbolic music. But you got enough data, you throw a deep enough network to it, and you have enough compute power, and you have the tools necessary to put that on YouTube, you can do pretty well.

Peter: A term I hadn't heard before there. Did you say math rock?

Bob: Math rock.

Peter: Which is?

Bob: Oh, it's a great, great genre. I mean, it's really complicated. It sounds complicated. Rock music. I mean, it requires a lot of coordination from humans to play this.

Peter: What characterizes it?

Bob: Let's see. Well, you just have to listen to it. A lot of people find it so complex, that's why they call it math rock.

Vincent: I think Hella is a good example. I really like that.

Peter: Okay. Hella. So this is reminding me of transformers here and what they do to language in that by adding more data, more parameters, they step up from being something that is like word completion on your smartphone, to handling paragraphs at a time of context and where it can make sense. You can give it a theme, “Write something about this topic,” and it will generate three paragraphs in a logical[ly] ordered progression that sound like someone thought about it. Is that what’s happening here? Only you haven’t, I think, said that you’re using transformers, but is it something equivalent?

Vincent: Yeah, in this case, the model that I mentioned is a recurrent neural network. So it’s also a sequence model that takes in some sequence and then can also generate some sequence. So you can feed it some something like a starting sequence, and it will try to, in the same sense as word completion, but then try to be sort of like a paragraph completion, sort of like, continue this audio part.

Peter: Right. I think one of Google’s chatbots early on was called Sequence-to-Sequence. So when people from outside this field come in and visit, and they talk about creativity, is this the big question as to what AI music is revealing about creativity in humans? Is that a conversation that comes up a lot? Because it sure seems to show up in the abstracts of a number of papers.

Bob: I think so. I would say in my work with folk music and Irish traditional music, and particularly applying algorithms to generate more of the same, I find a lot of people recoil. Well, not a lot, some people recoil in horror, because the idea is that music comes from the soul, and how dare you fiddle with Irish traditional music that is of the lands and hundreds of years and a whole community of people are safeguarding its safe passage to the future generations, and here you come with a computer to say, “Hey, look what it can do, it can produce more of the same.” And it is a challenge to a myth about the origins of music and what inspires people to create. But to my own work, in a sense, I’ve cursed myself, because the systems that I’ve developed have generated hundreds of thousands of tunes, and many of them are really good, and I want to work with many of them. And so now my life’s mission is to bring life to as many as I can before I die. And of course, creating a new model, we’ve just recently applied transformers to this kind of data. We don’t need billions of parameters, we just need a little over a million parameters and we’re creating results of the same quality as the earlier models of folk-rnn, which had 5 million parameters. But I need to be careful looking at new output of our system because I find stuff that I want to work with again, and I’ve got a growing repository of these things. But okay, back to the notion of creativity. An important thing that’s missing is the ability of a system to reflect and judge its creation and be able to revise. And a lot of these systems are just spitting out token after token, “What comes next in the sequence?”, rather than thinking about, “Okay, how can I revise what I’ve written in order to meet the challenge of this composition?” So much of music composition is problem-solving - coming up with a melody, fitting lyrics to a melody, making a melody memorable, orchestrating, arranging and developing, and so on.

Peter: Which prompts me to realize that I have not talked about using AI to analyze music. And I have a note here that OpenAI developed MuseNet, a GPT-2 backed AI that can generate compositions where you have sliders and parameters where you can say different styles. So do something like play Chopin in the style of Bon Jovi and things like that. But then it generated this spatial map, showing which composers it thought were musically similar. But some of the relationships were a bit odd, like it put Pachelbel next to Wagner. Is there a field there of using AI to analyze music to tell us things about it that maybe we didn't realize before or couldn't know because it required too much computation or memory?

Dorrien: There's definitely a big field. And a lot of people are doing sort of analysis of music. I think maybe Vincent has a little bit more experience with this. I think the group of Anya does quite some things.

Vincent: Yeah. So I think maybe it's because that's where I started my research. But my feeling is that that's actually a bigger part of the AI and music intersection, where we use AI to look at large data sets of music by different composers, for example. And you could say something interesting about what's the difference between this composer's style from that composer's style? And, of course, there are musicologists doing that. They're doing that because they know a lot about that music. But I think AI, or data science, or whatever you want to call it, in that sense, can show interesting details or other kinds of perspectives on how these genres or these composers differ, for example. That's one way how you could use these tools. And of course, then, you can use again for some kind of creative path. So if you want to write music, you can learn something there to write music.

Peter: Could it be used in other ways, like, for instance, detecting plagiarism and saying, "Look, this piece of music over here was written for the guitar, but it's a straightforward transformation of this organ piece that someone else wrote"?

Vincent: So in the AI Song Contest, there were some teams who, when they trained their models on a data set, they checked if the output they generated was part of the input. So what they did is they took the output and just looked at all the little parts in the output, basically, and checked whether those parts may appear in the input data set. That's one way of using [it]. So that's not on the creative side there, but it's checking for plagiarism, whether whatever you created as output might infringe on some rights.

Bob: Do you remember Big Data? Remember the term Big Data? That was a hot topic 10 years ago. And one of the movers and shakers of applying big data methods to the study of music is David Huron, who has, in some sense, controversy around this idea of assembling large collections of music to study the population. Much of musicology is about the use of music and its function in society. Music analysis, on the other hand, looks at particular pieces or particular practices, or composers to figure out how things work, how things are constructed. But big data and music is still a contentious topic. And when you bring in artificial intelligence and the systems that you don't know what they've learned, into the mix, it creates even more contentiousness. Now, back to this Pachelbel being close to Wagner. How could this be? They're separated by hundreds of years. But these are composers of some of the most played pieces at weddings. And so perhaps in that data set that they used, they found, on playlists of weddings, Pachelbel's "Canon in D" and Wagner's "Bridal March" often appear. So this would tell a computer, "Ah, these are similar because they're on the same playlist."

Peter: Thank you. That hadn't occurred to me. That's the equivalent of AI seeing sheep in a landscape where there aren't any, but because sheep are usually in landscapes, it's learned an association that isn't the one that I was thinking of when I look at that diagram. I'm aware we're getting close to time here. So let's go around here and say where do you think or where would you like this field to be 10 years from now. Dorrien?

Dorrien: Well, that's a tricky question because if we totally solve AI music creation, then all of us will be out of a job. So I think if we have powerful models, if we have great data sets, then we can do lots of cool things with that, we can create a lot of models and focus on some of the other aspects like production, to really come to models that people will have on their phone and listen to every day. That would be an amazing thing to see.

Peter: Thank you. Vincent.

Vincent: What I'd like to see is models becoming more useful for musicians so that they're using this technology to basically get their creativity to a different level, find new ways to be creative. And I think this is happening. And I also think it will help with democratization of music-making in a way. So that kind of ties into what you were saying about what people are afraid of when they think about automating or AI and music, is this sort of idea that there's some genius inspiration needed that's only accessible for a certain number of people in the world. This is very European-centric, or a Western way of thinking about music. I think there are a lot of other musical traditions in the world that view this very differently, more as a shared practice. And I think AI and these kind of tools can help with that. Get these in the hands of more people and let them express themselves through music. I think that would be an awesome future.

Peter: Thank you. Bob.

Bob: Ten years from now, I hope that the community continues to develop code that's open and accessible to a variety of people. You don't have to be particularly intelligent in order to do work in this area. You have to have some chops reading code and being able to download it and use it, but after that, I mean, it's amazing how much progress is made simply because people make their code available. The second thing I would like to see dealt with are economic questions. There's serious problems with the consolidation of media in a few players. And with these automated systems, there's going to be continued decrease in the cost of labor and people are going to find it even harder to make a living at these tasks that they used to be able to do that are now automated. So I really think there needs to be serious discussion on economic and ethical issues of the application of AI to music, to design, to many domains of life.

Peter: Wow. Thank you very much. Dorrien Herremans, Vincent Koops, Bob Sturm, thank you for coming on *AI and You*.

That's the end of the interview. I found it fascinating, I know I learned a lot.

In today's news ripped from the headlines about AI, New Scientist reports in an arresting headline that human brain cells in a dish learned how to play Pong faster than an AI. I know, many questions, right? Plus, this sounds super dystopian. Now they're growing human brains so they can enslave them to play primitive video games? Move over, George Orwell. Let's see if the truth is less alarming. There's a quote from the lead researcher, "We think it's fair to call them cyborg brains." Erm, not getting much comfort yet. And two days earlier, New Scientist ran an article titled "Crude eyes form on brain blob in a dish," about how "lumps of neural tissue were coaxed into sprouting rudimentary eyes." Okay, there's definitely more research going on into brain cells in dishes than I had previously suspected. But anyway, the research uses only a few hundred thousand human neurons, grown from scratch, not harvested from grad students, and it turns out while the research, at Cortical Labs, led by Brett Kagan, is able to teach the cells, which they call, no kidding, *DishBrain,* to play Pong faster than we can teach AI the game, because they were showing learning within 5 minutes, the cells aren't as good at it. Also the cells may be of "human or rodent origin." They know what the game is through electrical stimulation that encodes the parameters of the game. Score one for neuroplasticity, I guess.

Next week, I will be talking with David Danks, professor of Data Science & Philosophy at the University of California, San Diego, who lives at the intersection of cognitive science, AI, and ethics, and has developed an architecture for cognition.

That's next week on *AI and You*. Until then, remember: no matter how much computers learn how to do, it's how we come together as *humans* that matters.

http://aiandyou.net